



University
of Glasgow

Adamu-Lema, Fikru (2005) Scaling and intrinsic parameter fluctuations in nanoCMOS devices. PhD thesis.

<http://theses.gla.ac.uk/7086/>

Copyright and moral rights for this thesis are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

UNIVERSITY
of
GLASGOW

Thesis

**Scaling and Intrinsic Parameter Fluctuations in nano-
CMOS Devices**

by

Fikru Adamu-Lema

BEng Electronics, Middlesex University, London 1998

MRes., University College London, London 2000

Academic advisers

Prof. Asen Asenov, and

Dr. Scott Roy

Submitted to the University of Glasgow,
Department of Electronics and Electrical Engineering,
in full fulfilments of the requirements for the degree of
Doctor of Philosophy

21st day of September 2005

Abstract

The core of this thesis is a thorough investigation of the scaling properties of conventional nano-CMOS MOSFETs, their physical and operational limitations and intrinsic parameter fluctuations. To support this investigation a well calibrated 35 nm physical gate length real MOSFET fabricated by Toshiba was used as a reference transistor. Prior to the start of scaling to shorter channel lengths, the simulators were calibrated against the experimentally measured characteristics of the reference device. Comprehensive numerical simulators were then used for designing the next five generations of transistors that correspond to the technology nodes of the latest International Technology Roadmap for Semiconductors (ITRS).

The scaling of field effect transistors is one of the most widely studied concepts in semiconductor technology. The emphases of such studies have varied over the years, being dictated by the dominant issues faced by the microelectronics industry. The research presented in this thesis is focused on the present state of the scaling of conventional MOSFETs and its projections during the next 15 years.

The electrical properties of conventional MOSFETs; threshold voltage (V_T), subthreshold slope (S) and *on-off* currents (I_{on} , I_{off}), which are scaled to channel lengths of 35, 25, 18, 13, and 9 nm have been investigated. In addition, the channel doping profile and the corresponding carrier mobility in each generation of transistors have also been studied and compared. The concern of limited solid solubility of dopants in silicon is also addressed along with the problem of high channel doping concentrations in scaled devices.

The other important issue associated with the scaling of conventional MOSFETs are the intrinsic parameter fluctuations (IPF) due to discrete random dopants in the inversion layer and the effects of gate Line Edge Roughness (LER). The variations of the three important MOSFET parameters (I_{off} , V_T and I_{on}), induced by random discrete dopants and LER have been comprehensively studied in the thesis.

Finally, one of the promising emerging CMOS transistor architectures, the Ultra Thin Body (UTB) SOI MOSFET, which is expected to replace the conventional MOSFET, has been investigated from the scaling point of view.

Acknowledgements

It was a great privilege to be a member of device modelling group at the University of Glasgow, where the work presented in this thesis was conducted during the last 4 years. There are many academics who have helped, inspired and motivated me throughout these challenging years. Their support was an indispensable instrument to accomplish this research programme.

Most of all, I would like to thank Professor Asen Asenov, my principal academic adviser. This thesis would not have been possible without his unreserved support. I am grateful for his precise guidance, countless enlightening discussions, and corrections during my research and writing-up of this thesis. His deep knowledge and enthusiasms for semiconductor physics and devices is one of the many motivating factors that kept me going during difficult times.

I would also like to thank Dr. Scott Roy, my second academic adviser, for all his help. In particular, I am grateful for his invaluable comments, corrections, and his eagerness to help and willingness to share his vast knowledge. His approachability, resourcefulness, and intellect are an inspiration and great assets to the device modelling group.

I have gained enormous knowledge over the years, in all aspects of life and professional skills from the device modelling group members to whom I am greatly indebted for their cooperation and support. Above all, it was great fun to work with my officemates, Craig Alexander and Gareth Roy, in the “famous” IBM room. Thanks also go to Richard Wilkins for helpful discussions, proof reading of my thesis and useful feedback. I am also grateful to Andrew Brown, Binjie Cheng, and Karol Kalna for their supports and encouragements.

Without the patience and help of the department’s IT support group, most simulation work and data processing jobs would have been impossible. Thanks to Billy Allen and Jon Trinder for facilitating an efficient computer network and timely maintenances from which I have benefited a lot.

Finally, I would like to thank IBM for facilitating this research program and the Toshiba Research and Development Centre for supplying us with the experimental data of their prototype 35 nm CMOS device.

List of tables

Table 2:1	The near term years (2003-2009) of selected overall roadmap technology characteristics that are required to maintain the present scaling trends of conventional MOSFETs. <i>hp</i> stands for Half Pitch..	13
Table 2.2	The long - term years (2010-2018).....	14
Table 2:4	Summary of the constant field scaling and the generalised scaling rules.....	18
Table 3.1	Fitting parameters of concentration dependant mobility model.....	71
Table 4:1	Important device dimensions and doping information of the 35 nm gate length <i>p</i> -type MOSFET.....	86
Table 4:2	Electrical parameters of the realistic Toshiba 35 nm gate length <i>n</i> -type MOSFET.....	86
Table 4:3	Initial estimation of numerical values obtained from the curve fitting of the two equations (4.1a, b) to the experimental data of the Toshiba device. The fitted curve is illustrated in figure 4.3 with the highlight of the projection range.....	91
Table 4:4	Summary of the electrical parameters for the calibrated n and p-type 35nm model MOSFET.....	113
Table 4:5	Summary of the important device dimensions and impurity doping values of the calibrated n and p-channel of the calibrated 35 nm y MOSFETs.....	113
Table 4:6	Near term predictions of technology road map for semiconductors: (ITRS 2003 edition).....	115
Table 4:7	Long term predictions of technology road map for semiconductors (ITRS 2003 edition).....	115
Table 4:8	Comparative values of supply voltage, drive current and off current for the reference 35nm and scaled 25, 18, 13 and 9nm devices.....	120

Table 4:9	Average channel mobility values in the reference and scaled devices at the corresponding supply voltages.....	120
Table 5:1	Summary of selected statistical parameters.....	146
Table 5:2	ITRS guidelines for LER, LWR, and printed gate CD variation.....	152
Table 5:3	Comparison of electrical parameters 35 and 25 nm gate length of <i>n</i> -channel MOSFETs with LER effect.....	160
Table 5:4	Summaries of device parameters with the effect of LER.....	160
Table 5:5	<i>Off</i> -current and <i>on</i> -current in 35 nm <i>n</i> -MOSFET with LER of different correlation lengths.....	164
Table 6:1	ITRS requirements of the silicon layer and SiO ₂ thickness in 2001 and 2003 editions.....	177
Table :6:2	Summaries of investigated scaling scenarios UTB-SOI MOSFET....	179
Table 6:3	Investigations of various scaling scenarios.....	170
Table 6:4	Carrier mobility in UTB SOI and scaled conventional MOSFETs....	185

List of figures

Figure 2:1	Visualization of Moore’s Law: The number of transistors integrated in a commercially available processor and the outlook towards a billion transistors in a single processor due in year 2007 (Intel). Inset: Moore’s original projection [2.1].	9
Figure 2:2	Cost and functionality (data from ITRS 2004 update).....	10
Figure 2:3	The technology trends of on-chip local clock, off-chip frequency, chip density in SRAM and Logic gates (transistors per cm2). ITRS 2004 Edition.....	11
Figure 2:4	Technology half-pitch and gate length trends.....	16
Figure 2:5	Illustration of MOSFET miniaturisation. The sketch on the right hand is the scaled device according to the constant field rule. (Reference [2.15]).....	17
Figure 2:6	Cross-sections of fabricated CMOS device and an inverter circuit that can be realized.....	22
Figure 2:7	Cross section view of LDD <i>n</i> -MOSFET (A) and MOSFET with heavily doped source and drain extensions.....	23
Figure 2:8	schematic illustrations showing self-aligned silicide in source/drain and on the poly-gate, halo (super-halo) doping and retrograde channel doping in a cross sectional view of the MOSFET. The inset depicts the ideal low-high (retrograde) doping profile.....	24
Figure 2:9	Visual illustrations of quantum effects near the Si/SiO2 interface. Reference [2.42].....	27
Figure 2:10	Junction leakage current due to band to band tunnelling (reference [2.44]).....	29
Figure 2:11	The overall gate tunnelling current is the sum all the components tunnelling current namely the source-to-gate, drain-to-gate and channel to gate currents (reference [2.46])..	30

Figure 2:12	The relationship between the gate leakage current and the physical gate length. Inset: the gate current density for various oxide thicknesses as a function of gate voltage. Data from ITRS (2004 up-date) and the inset graph is adopted from reference. [2.48].....	31
Figure 2:13	Total power dissipation as a function of technology node and an illustration of the increasing of the static power dissipation due to the significant increase in the off-state current as technology scales to 65 nm and beyond. Data source: ITRS 2004 and references [2.64] [2.65].....	35
Figure 2:14	Schematic illustration of the relationship between the defect generation rate, defect density, oxide thickness and the stress voltage: Reference [2.70].....	38
Figure 3:1	Flow chart that illustrate the integrated process and device simulation methodology.....	51
Figure 3:2	Summary of CMOS device fabrication steps	53
Figure 3:3	A three dimensional schematic visualisation of an average ion's projected range and its distribution in an idealised semiconductor material.....	57
Figure 3:4	Discrete dopant distribution imported from DADOS in to the 35nm test MOSFET.	63
Figure 3:5	Illustration of the oxidation process adapted from [3.36].....	64
Figure 3:6	Hierarchy of the semiconductor device simulation models, Ravaoli [3.47].....	68
Figure 3:7	Concentration dependant mobility model with various values of α_n comparing with the default values given in the Taurus device simulator. The star and open square represents calculated values of mobility using α_n values of 0.7 and 0.65 in equation (3.15) and the continues line denote the mobility values as a function of carrier concentration specified in Medici device simulator.	72

Figure 4:1	Flow chart that illustrates the systematic calibration methodology.....	85
Figure 4:2	Transmission electron microgram (TEM) photograph of the 35nm <i>n</i> -channel device from which all data and process information have been obtained for this work. Ref. [4.3]	85
Figure 4:3	Two Gaussians are illustrated with their corresponding projection range, (R_{p1} and R_{p2}) and standard deviations (σ_1 and σ_2). These two curves represent the two arbitrary Gaussians distributions, given by equations 4.1 and 4.2. The point of this figure is to illustrate the practical approach, which highlights that the combined effect two virtual implantation may yield the final doping profile of the channel implantation	89
Figure 4:4	The sum of the two Gaussian equations in (4.1, 4.2) is fitted to the experimental data in order to approximate the combined projection range and standard deviation. The symbols represent the measured Toshiba data and the line represents the sum of the two equations. The approximated values obtained from this fitted curve are given in table 4:1....	90
Figure 4:5	The resultant (calibrated) channel doping profile of <i>n</i> -channel MOSFET, approximated from the two Gaussian distributions given by equations 4.1 and 4.2.....	93
Figure 4:6	The calibrated doping profiles of the channel (indium dopants) and the source and drain extensions (arsenic dopants). The square and star symbols represent the measured data of indium and arsenic respectively. The corresponding simulation results are indicated by open circle and '*'.....	94
Figure 4:7	2-D illustration of continues doping of the channel profile of the <i>p</i> -channel MOSFET (a), 1-D doping profile of As in the lateral direction (cross-section BB) and the channel doping profile in the transverse direction (cross-section CC).....	96
Figure 4:8	Halo doping profile of the calibrated 35 nm <i>n</i> -channel MOSFET.....	97

Figure 4:9	Relation of dielectric constant of silicon oxynitride to the mole fraction of silicon dioxide and silicon nitride calculated using the Clausius-Mossotti equation.....	101
Figure 4:10	2D electric potential in the simulated 35 nm <i>n</i> -channel MOSFET at $V_G = 0V$ and.....	103
Figure 4:11	Threshold voltage estimation using the current criteria.....	104
Figure 4:12	Well calibrated I_d - V_g characteristics of 35 nm <i>n</i> -channel MOSFET at a high drain voltage of 850mV.....	106
Figure 4:13	Well calibrated I_d - V_g characteristics of the 35 nm <i>n</i> -channel MOSFET at a low drain voltage of 50mV.....	106
Figure 4:14	Output characteristics of the calibrated 35 nm <i>n</i> -channel MOSFET (I_d - V_d).....	107
Figure 4:15	I_d - V_g characteristics of the calibrated 35 nm <i>p</i> -channel MOSFET on a logarithmic scale	108
Figure 4:16	I_d - V_g characteristics of the calibrated 35 nm <i>p</i> -channel MOSFET at low V_d on linear scale.....	108
Figure 4:17	Output characteristics of the calibrated 35 nm <i>p</i> -channel MOSFET (I_d - V_d).....	108
Figure 4:18	The effect of different mobility models on the carrier mobility in the 35nm prototype transistor.....	109
Figure 4:19	Carrier (electron) mobility versus the lateral distance across the channel at a low drain voltage of 50mV. The dashed lines in the vertical direction indicate where the gate edges are located.....	110
Figure 4:20	Carrier (electron) mobility versus the lateral distance across the channel at the high drain voltage of 850mV. The dashed lines in the vertical direction indicate where the gate edges are located.....	111
Figure 4:21	2D net doping profile and important design parameters of the calibrated 35 nm gate length model.	112

Figure 4:22	Doping profiles of the channel region in the reference 35 nm and the scaled devices of 25, 18, 13, and 9 nm gate length. The dashed lines show the channel doping profile of the scaled devices. The solid triangle and star symbols represent the experimental data for indium channel doping and arsenic doping in extensions respectively. The background doping ($\sim 3 \times 10^{17}$) is marked by the horizontal dashed line.....	118
Figure 4:23	I_d - V_g characteristics of the scaled 25 nm MOSFET (a) and its overall output characteristics (I_d - V_d) The dotted curve on I_d - V_d characteristics corresponds to the gate voltage which is equivalent to the supply voltage of 800mV.....	121
Figure 4:24	I_d - V_g characteristics of the scaled 18 nm MOSFET (a) and its overall output characteristics (I_d - V_d) The dotted curve on I_d - V_d characteristics corresponds to the gate voltage which is equivalent to supply voltage of 700mV.....	122
Figure 4:25	I_d - V_g characteristics of the scaled 13 nm MOSFET (a) and its overall output characteristics (I_d - V_d) The dotted curve on I_d - V_d characteristics corresponds to a gate voltage, which is equivalent to a supply voltage of 600mV.....	123
Figure 4:26	I_d - V_g characteristics of the scaled 9 nm MOSFET (a) and its overall output characteristics (I_d - V_d) The dotted curve on I_d - V_d characteristics corresponds to a gate voltage which is equivalent to a supply voltage of 500mV.....	124
Figure 4:27	Drain current as a function of gate voltage on both a logarithmic and linear scale to highlight the electrostatic properties and drive current in the saturated regime respectively. The logarithmic scale includes the experimental data for the reference 35 nm Toshiba MOSFET and calibrated device.....	125
Figure 4:28	Comparison of the gate length with the off and on-currents of MOSFETs from reported data, ITRS'03 requirement trends and simulation results of this work (a) and the corresponding	

	drain current (b).....	126
Figure 4:29	The electric field profile across the channel just below the Si/SiO ₂ interface. The peak points in each device occur near the drain edge of the <i>p-n</i> junction.....	127
Figure 4:30	Net doping concentration in the channels of the reference 35nm and the scaled 25, 18, 13, and 9 nm gate length devices. The concentration points indicated by the arrows are located in the geometric centre of the corresponding devices.....	128
Figure 5:1	The evolution of atomistic devices: (a) MOSFET with continuous ionised dopant charge and smooth boundaries and Si/SiO ₂ interfaces; (b) Sketch of 22 nm MOSFET required for 45 nm technology node with rough interface, LER and random discrete dopants; (c) Impression of 5 nm MOSFET with the silicon crystal lattice superimposed. Courtesy of Asenov et al.[5:1].....	135
Figure 5:2	Image of 3D potential distribution in one of the simulated 35 nm MOSFETs illustrating the position of discrete random dopants in the channel and source/drain regions.....	136
Figure 5:3	Spread Id-Vg characteristics of 200 transistors with 35 nm gate length.....	138
Figure 5:4	Threshold voltage variation as a function of scaled device gate length.....	139
Figure 5:5	Histograms of VT distributions for the prototype 35 nm and the scaled 25 and 18 nm MOSFETs.....	141
Figure 5:6	Mean and standard deviation of I _{off} in the 35 25 and 18 nm MOSFETs.....	142
Figure 5:7	I _{off} distribution of the calibrated and scaled (25 and 18nm) devices both in linear scale and logarithmic scale.....	143
Figure 5:8	3 rd and 4 th moments of the I _{off} distribution in 18, 25, and 35 nm MOSFETs.....	144
Figure 5:9	Standard deviation and mean of drain current in the strong inversion region.....	145

Figure 5:10	Normalized histogram of drain current distribution in the strong inversion region of the 35 nm; the 25 nm and 18 nm <i>n</i> -channel transistors.....	147
Figure 5:11	Circuit schematics of SRAM cell (a); Distribution of SNM; adapted from [5.].....	148
Figure 5:12	Comparative illustrations of (a) an ideal transistor with straight gate edges and uniform channel width and (b) a top view of hypothetical gate edges which are constructed from a pair of the randomly generated rough lines used in this work for to investigate LER.....	149
Figure 5:13	I_d - V_g characteristics for a sample of 50 <i>n</i> -channel MOSFETs Including LER effect with RMS height of 1.nm. Values at high drain voltage of 850mV are plotted in logarithmic scale and those at low drain voltage of 50mV are plotted in linear scale. The inset is net doping profile of two samples that shows the minimum and maximum off currents.....	154
Figure 5:14	1-D net-doping profile in the interface and mobility of device A at the cross section A-A.....	155
Figure 5:15	1-D net doping profile in the interface and mobility of device B at the cross section A-A.....	155
Figure 5:16	I_d - V_g characteristics of 50 <i>n</i> -channel MOSFETs, RMS height of 2 nm.....	155
Figure 5:17	I_d - V_g characteristics of 50 <i>n</i> -channel MOSFETs including the LER effects with the RMS height of 3 nm. The inset image depicts the two extreme cases happed as a result of the degree of LER in devices E and F among the simulated 50 devices...	156
Figure 5:18	LER with different values of Δ	157
Figure 5:19	1-D channel doping profiles near the interface of devices shown in row three (filled symbol) and row four (open symbols) in figure 5:13.....	158
Figure 5:20	Histogram showing distributions of intrinsic device	169

parameters in the 35nm n-MOSFET. The columns and rows represent RMS values and the corresponding distribution of electrical parameters respectively. The correlation length is 30nm in all cases. Row 1 and 2 show linear scale and logarithmic scale of the *off*-current distributions respectively..

Figure 5:21	Plot of linear scale histogram of I_{off} distribution in the 25 nm transistor with RMS=3 nm.....	161
Figure 5:22	Plot of logarithmic scale of histogram of I_{off} distribution in the 25 nm transistor with RMS=3 nm.....	161
Figure 5:23	Histogram of drive current distribution for 25 nm transistor with RMS =3 nm.....	161
Figure 5:24	Gate shape and arsenic doping profile of three randomly generated MOSFETs using correlation lengths 30, 15, and 5 nm. Row 1 shows top view of gate shape before As-implantation and rows 2-3 depict the mapping of arsenic concentration, 5 seconds and then after 15 seconds annealing respectively and row 4 shows distribution of arsenic and <i>p-n</i> junctions after full process.....	162
Figure 5:25	I_d - V_g Characteristics at high drain voltage of 850mV for the 35 nm n-channel MOSFET (a-log scale and b-linear scale) with the effect of LER simulated for various correlation lengths.....	163
Figure 5:26	2-D doping profiles of arsenic dopants (first column) and net doping profiles (second column) in different positions along the device width. The figures in rows correspond to the doping profiles in the marked cut planes position (a), <i>p-n</i> junction (b), and the 3-D doping profile of arsenic is shown in (c).....	166
Figure 5:27	1-D doping profiles in the Si/SiO ₂ interface along the channel of the same device but at different position along the gate width as shown in figure 5:19a.....	167
Figure 5:28	Basic geometric elements of LER.....	168

Figure 5:29	I_d - V_g characteristics of the basic geometric gate patterns shown in figure 5:23.....	168
Figure 6:1	Typical PDSOI (a) and FDSOI (b) structures.....	175
Figure 6:2	Generic diagram of ultra-thin SOI used for investigations of various scaling schemes in this work.....	179
Figure 6:3	linear I_d - V_g characteristics at low drain voltage of $V_{dd} = 50$ mV of the same devices.....	181
Figure 6:4	I_d - V_g characteristics of the UTB-SOI with $L_c = 15, 10$ and 5 nm.....	182
Figure 6:5	Critical performance parameters.....	183
Figure 6:6	V_T , ΔV_T and subthreshold slope.....	183
Figure 6:7	I_d - V_d (a) characteristics of the 15 nm channel.....	183
Figure 6:8	I_d - V_d (a) characteristics of the 10 nm channel.....	184
Figure 6:9	I_d - V_d (a) (b) characteristics of the 5 nm channel.....	184
Figure 6:10	Electric field (a), electron velocity (b) and electron mobility of the selected UTBSOI MOSFETs with the channel lengths of $L_c=15, 10$ and 5 nm.....	186

List of symbols

Symbol	Description	unit
α	Scaling factor for electric field	-
β_w	Weibull slope	-
Δ	Root mean value (RMS)	-
ϵ_{si}	Silicon permittivity	F/cm
ϵ_{ox}	Oxide permittivity	F/cm
κ	Scaling factor for device dimensions	-
Λ	Correlation length	
μ_e	Electron mobility	cm ² /Vs
μ_p	Hole mobility	cm ² /Vs
μ_{min}	Minimum electron/hole mobility	cm ² /Vs
μ_{max}	Maximum electron mobility	cm ² /Vs
ρ	Charge density	C/cm ³
τ_{int}	Inverter delay time	[s]
ψ	Potential	V
ψ_{bi}	Built in potential	V
A	Area	cm ²
C_{gate}	Gate capacitance	F
C_{in}	Input capacitance	F
C_{out}	Output capacitance	F
E	Electric field	V/cm
E_F	Fermi level	J
E_g	Energy gap	J
G_n	Electron generation rate	cm ⁻³ s ⁻¹
G_p	Hole generation rate	cm ⁻³ s ⁻¹
g	Density of substrate semiconductor material	gm/cm ³
h	Planck's constant (6.625×10 ⁻³⁴ J-s)	Js
\hbar	Reduced Planck's constant ($\hbar = h/2\pi$)	Js
I_{ds}	Fringing current from drain to gate	A
I_{dsat}	Saturation current (drain)	A
I_g	Gate leakage current	A
I_{gc}	Tunnelling current from channel to gate	A
I_{gs}	Fringing current from source to gate	A
I_{off}	Off current	A
I_{on}	On current	A
J_{B-B}	Band to band tunnelling current density	A/cm ²
J_{DT}	Gate tunnelling current density	A/cm ²
J_n	Electron current density	A/cm ²

J_p	Hole current density	A/cm ²
k	Boltzmann's constant (1.38×10^{-23})	J/K
L_{ch}	Channel length	nm
L_g	Gate length	nm
L_{el}	Effective channel length	nm
m^*	Effective electron mass	kg
M_i	Atomic mass of implanted atom	gm
M_s	Atomic mass semiconductor	gm
N_a	Acceptor doping concentration	/cm ³
N_{BD}	Defect density to break down	-
N_d	Donor doping concentration	/cm ³
P	Power	W
P_D	Dynamic (active) power dissipation	W
P_{gen}	Defect generation rate	-
P_{sh}	Power dissipation due to occurrence of short circuit	W
P_{st}	Static power dissipation	W
P_T	Total power dissipation	W
q	Electron charge (1.6×10^{-19} C)	C
Q_{BD}	Charge to break down	C
Q_{inj}	Injection of electron charge	C
r	Lag number (in autocorrelation function)	C
R	Range	nm
R_p	Projection range	nm
R_{sw}	Switching resistance	Ω
S	Subthreshold slope	V/decade
S_e	Electron stopping power	W
S_n	Nuclear stopping power	W
t_{BD}	Time to break down	s
t_{ox}	Oxide thickness	nm
V_{dd}	Power supply voltage	V
V_T	Threshold voltage	V
ΔV_t	Threshold voltage deference	V
W	Device width	nm
W_{dm}	Drain junction depletion width	nm
W_s	Source junction depletion width	nm
Δx	Regular spacing of samples in autocorrelation function	
x_j	Source/Drain junction depth	nm
Z_i	Atomic number of implanted atom	-
Z_s	Atomic number of semiconductor material	-

Table of Contents

ABSTRACTii

ACKNOWLEDGMENTiii

LIST OF TABLES.....iv

LIST OF FIGURES.....vi

LIST OF SYMBOLSxiv

1

INTRODUCTION.....1

1.1 MOTIVATION AND DESCRIPTION OF THE PROBLEMS.....2

1.2 RESEARCH OBJECTIVES.....3

1.3 THESIS OUTLINE.....4

1.4 CHAPTER REFERENCE..... 6

2

THE SCALING OF MOSFETS, MOORE’S LAW, AND ITRS 7

2.1 THE IMPACT OF MOORE’S LAW AND ITRS ON DEVICE SCALING.....8

2.1.1 *Moore’s Law* 8

2.1.2 *Implication of Moore’s law*..... 10

2.1.3 *The International Technology Roadmap for Semiconductors* 12

2.1.4 *The scaling factors and technology trends* 14

2.2 THE SCALING RULES FOR CONVENTIONAL MOSFETS.....17

2.2.1 *Constant field scaling* 17

2.2.2 *Generalised scaling rule*..... 19

2.2.3 *Evolution of CMOS design* 22

2.2.4 *The present state and the future trends*..... 25

2.3 FACTORS LIMITING THE SCALING OF CONVENTIONAL MOSFET.....26

2.3.1 *Quantum mechanical tunnelling*..... 27

2.3.1.1 *Band to Band tunnelling* 28

2.3.1.2	Direct gate oxide tunnelling.....	29
2.3.1.3	Source to drain tunnelling.....	32
2.3.2	<i>Intrinsic parameter fluctuations</i>	32
2.3.3	<i>Power dissipation</i>	34
2.3.4	<i>Reliability of the Ultrathin Gate Oxide</i>	36
2.4	CHAPTER SUMMARY.....	39
2.5	CHAPTER REFERENCE.....	40
3		
SIMULATION METHODOLOGY.....		47
3.1	TCAD AND INTEGRATED PROCESS & DEVICE SIMULATION.....	47
3.1.1	<i>The role of TCAD in advanced device design and characterisation</i>	48
3.1.2	<i>Integrated process and device simulation in 3D</i>	50
3.2	FRONT END PROCESS SIMULATION – 3D.....	52
3.2.1	<i>Implantation – The Introduction of impurities in to Si crystal</i>	54
3.2.2	<i>Ion stopping energy and the projection range</i>	56
3.2.3	<i>The Gaussian and dual Pearson implantation models</i>	58
3.2.4	<i>Diffusion</i>	61
3.2.5	<i>Diffusion and Defects, Object-oriented Simulator (DADOS)</i>	62
3.2.6	<i>Oxidation</i>	63
3.2.7	<i>Gate dielectric material</i>	65
3.3	DEVICE SIMULATION.....	66
3.3.1	<i>Equations governing device operation</i>	66
3.3.2	<i>Comparisons of the diverse approaches to transport simulation</i>	67
3.3.3	<i>Implemented physical models</i>	70
3.3.2.1	Concentration dependence.....	71
3.3.2.2	Perpendicular Electric-Field Dependence	73
3.3.2.3	Parallel Electric-Field Dependant Mobility Model:	73
3.3.2.4	Quantum Mechanical Effect - MLDG Approximation.....	74
3.3.2.5	Glasgow Atomistic Device Simulator: 3-D.....	74
3.4	CHAPTER SUMMARY.....	75
3.5	CHAPTER 3 REFERENCES.....	76

SCALING OF THE CONVENTIONAL MOSFET DEVICES.....81

4.1	CALIBRATION STRATEGY AND ITS ROLE IN DEVICE SIMULATION.....	81
4.1.1	<i>Systematic calibration.....</i>	83
4.2	THE STRUCTURE OF THE REAL 35NM GATE LENGTH MOSFET.....	85
4.2.1	<i>The Doping profiles of the channel, and the source and drain extensions.</i>	87
4.2.1.1	Initial estimation of implantation parameters	88
4.2.1.2	Analytical calculation of R_p , σ_p and Φ of the two Gaussians	91
4.2.1.3	Calibrated channel doping profile: n-MOSFET	93
4.2.1.4	Calibrated doping profile: p-MOSFET.....	95
4.2.2	<i>The halo doping and well implantation of the n-channel MOSFET.....</i>	96
4.3	SiO_xN_y AS A DIELECTRIC MATERIAL.....	98
4.3.1	<i>Estimation of the SiO_xN_y dielectric constant</i>	99
4.4	CALIBRATION OF ELECTRICAL PARAMETERS AT LOW AND HIGH V_{DD}	102
4.4.1	<i>Simulation results of the well calibrated 35 nm n-MOSFET.....</i>	106
4.4.2	<i>Simulation results of the calibrated 35nm p-channel MOSFET.....</i>	107
4.5	MOBILITY AND ELECTRIC FIELD IN THE CALIBRATED 35 NM N-MOSFET...	109
4.6	DEVICE STRUCTURE AND PARAMETERS OF THE CALIBRATED 35NM MOSFET	112
4.7	SCALING STRATEGY.....	114
4.7.1	<i>The scaling scenarios of the conventional MOSFET devices.....</i>	114
4.7.2	<i>Doping in the scaled devices</i>	115
4.7.2.1	Scaling properties of R_p and σ	116
4.7.3	<i>Scaling results: Electrical properties of the scaled devices</i>	119
4.8	COMPARATIVE STUDY OF THE CALIBRATED AND SCALED DEVICE PROPERTIES	125
4.9	CONCLUSION.....	129
4.10	CHAPTER REFERENCES.....	130

INTRINSIC PARAMETER FLUCTUATIONS IN SCALED MOSFETS.....134

5.1	RANDOM DISCRETE DOPANTS.....	135
5.1.1	<i>Simulation Approach</i>	136
5.1.2	<i>Threshold Voltage variation.....</i>	139

5.1.3	<i>Off-current variation</i>	142
5.1.4	<i>Drain current variation in the strong inversion region</i>	145
5.1.5	<i>Impact of random discrete dopant fluctuation on circuits</i>	148
5.2	LINE EDGE ROUGHNESS IN CONVENTIONAL MOSFETs.....	149
5.2.1	<i>Contributing factors to LER</i>	150
5.2.2	<i>Characterisation of LER</i>	150
5.2.3	<i>LER Simulation approach</i>	152
5.2.4	<i>IPF due to LER</i>	153
5.2.4.1	The effect of LER on device parameters with various value of Δ	153
5.2.4.2	The effect of Λ	162
5.3	“DETERMINISTIC” LER SIMULATION	164
5.3.1	<i>Local doping variation</i>	165
5.3.2	<i>The effect Deterministic LER on I-V characteristics</i>	168
5.4	CHAPTER SUMMARY	169
5.5	CHAPTER REFERENCES	170

6

THE SCALING OF SILICON ON INSULATOR.....173

6.1	SOI DEVICE STRUCTURE.....	174
6.1.1	<i>Partially depleted and fully depleted SOI</i>	175
6.1.2	<i>Ultra thin body SOI</i>	177
6.2	SCALING STUDY OF UTB-SOI MOSFETs.....	177
6.2.1	<i>Electrical characteristics</i>	180
6.2.2	<i>Mobility and carrier velocity in UTB SOI</i>	185
6.3	CONCLUSIONS.....	187
6.4	CHAPTER REFERENCES.....	187

7

CONCLUSION AND SUGGESTIONS FOR FUTURE WORK..... 191

7.1	CONCLUSIONS.....	191
7.2	SUGGESTIONS FOR FUTURE RESEARCH.....	194

1

Introduction

Following their invention in 1947 at Bell Laboratories[†] [1.1], transistors have since dominated the semiconductor market as the main component of electronic circuits. It is not an exaggeration to claim that the introduction of the transistor to electronic circuits allowed today's microelectronics systems, computer technology, telecommunications, space exploration, and modern scientific research in general [1.2].

About a decade following the invention of the transistor, the development of integrated circuits (IC) opened up a new era for transistor applications in electronic circuits [1.3] [1.4] [1.5]. Since the first ICs, the market has continued to demand greater and greater functionality per package component. Industry is striving to satiate this demand by increasing the number of transistors on a chip. Naturally, however, at the same time industry insists on keeping the cost of production to a minimum.

The scaling down of transistors is widely regarded as the best way to meet the growing demand for a high volume of transistors integrated into a single chip. The idea of scaling conventional transistors [1.6], which is the main theme of this research, then gathered apace to revolutionize the advancement of the microelectronics industry.

This introduction discusses the motivation, problems and main objectives of the research presented in the thesis. It also outlines the overall structure of the thesis.

[†] Three scientists, William Shockley, Walter Brattain and John Bardeen, are credited for the invention of the transistor at Bell Laboratory in 1945 and all three received Nobel price for their contribution to physics.

1.1 Motivation and description of the problems

One of the important growth drivers for the semiconductor industry is the shrinking of transistors to smaller dimensions. The reduction of transistor size increases performance. System performance and functionality increases as the density of transistors which can be integrated onto a single chip increases. At the same time switching speed increases inversely proportionally with gate length (L_g), allowing faster circuit operation [1.7]. Another benefit of scaling transistors is the reduction in manufacturing costs per device.

Despite the challenges and difficulties regarding MOSFET miniaturisation, scaling still continues to dominate semiconductor research and development programmes. However, problems limiting the scaling of conventional MOSFETs continue to challenge. These limitations and the corresponding technological challenges are currently dictating a shift of research from conventional MOSFET to alternative devices.

According to the International Technology Roadmap for Semiconductors (ITRS), if Moore's law [1.8] is maintained the microelectronics industry expects the mass production of 7 nm physical gate length transistors in twelve to fifteen years' time. Realization of such atomic scale transistors is technologically challenging, if not unattainable. To reach this level of miniaturisation will require an enormous amount of research and resources. So far, the semiconductor industry has pursued a traditional performance trajectory [1.9]. However, keeping to this tradition by using conventional ways of scaling is a monumental task as we approach 'decananometre' (tens of nanometres) channel devices.

Unfortunately, the scaling of conventional MOSFETs is fast approaching its zenith [1.10]; there are some fundamental physical and material limitations that hinder the progress of scaling. Some of these fundamental problems are, for example: short channel effects, unacceptably high off-state currents which induce a high (static) power dissipation, lack of performance and intrinsic parameter fluctuations. The challenges of keeping to the present scaling trends for conventional MOSFETs and the increasing interest in alternative transistor architectures are the main motivations of this research work.

1.2 Research objectives

The first major objective of this research is by using commercial process and device Technology Computer Aided Design (TCAD) tools, to design well scaled conventional MOSFETs with gate lengths[‡] of 35, 25, 18, 13, and 9 nm, corresponding to the 90, 65, 45, 32, and 22 nm technology nodes respectively and to study in detail their electrical properties.

The scaling is based on a real state of the art 35 nm MOSFET, which has been used to guide the design of the scaled devices under study. The simulation tools have been appropriately calibrated in respect of this benchmark device. The study of the scaled devices aims to further enhance the qualitative and quantitative understanding of the main factors limiting the scaling of conventional CMOS in accordance with the requirements of the latest ITRS. A comprehensive investigation of the electrical characteristics and parameters of the scaled devices should lead to a more detailed understanding of the problems of scaling. The better understanding leads to possible solutions for the challenges which the semiconductor industry faces, up to and beyond the end of the ITRS in 2018. This work also tries to contribute answers to common questions asked by researchers and technologists, such as: “are we near the end of the ‘classic’ scaling pattern? When will it end and how? Can we go beyond ITRS’s present predictions of scaling? Do we really need new device architectures? If so how far can we go with them?”

The second major objective of this research is to investigate how far intrinsic parameter fluctuations affect the characteristics and the integration of the advanced MOSFETs. This was done by employing 3-D atomistic device simulation software

[‡] It is important to note that the gate length mentioned above and frequently used throughout this thesis correspond to the “physical gate” length rather than the “printed gate” lengths or channel length (effective channel length). According to the definitions given in the latest ITRS 2003 edition, printed gate length refers to the requirements by the semiconductor industry “as-printed” in photo resist prior to etching. On the other hand, the physical gate length (“as etched in polysilicon”) may be reduced from the “as-printed” dimension as a result of etching process. Since this work is mainly interested on high performance devices which derive the miniaturization of transistors, the physical gate length has been considered as a minimum feature size during the scaling.

[1.11] developed by the Device Modelling Group at the University of Glasgow. In essence, one is trying to answer the following questions: How can the effects of fluctuations be suppressed and by what means? Can we predict the extent to which intrinsic fluctuations hamper the progress of conventional MOSFET scaling? Based on the analysis of the information and data obtained from computer simulation and experimental results, an attempt has been made to find solutions to circumvent these problems.

1.2 Thesis outline

The thesis is divided into seven chapters of which three chapters (chapters 4, 5, and 6) contain simulation results and discussions. Chapters 2 and 3 contain back-ground information and a description of the simulation methodology developed and used to perform the research task. In the following sections, a condensed summary of each chapter is given.

Chapter 2 introduces Moore's law, the International Technology Roadmap for Semiconductors (ITRS) [1.12] and the scaling concepts of conventional MOSFETs. The emphasis is placed on the projected miniaturisation of MOSFETs according to the requirements and the technology nodes of the present ITRS edition. The chapter starts by reviewing Moore's law from a historical and economic perspective and its contribution to the advancement of the semiconductor industry. The discussion expands to the driving role of the ITRS in relation to device scaling and the complimentary relationships between Moore's law and the ITRS.

The two main approaches to scaling are then described, namely the constant field scaling and the generalised scaling rules. The fundamental limitations to conventional MOSFET scaling are also discussed in conjunction with these two approaches. The final section of Chapter 2 gives historical insights into the evolutionary development of field effect transistors (FETs) and looks at possible future directions, through a review of the present state of MOSFETs and the semiconductor industry.

Chapter 3 focuses mainly on the methods employed in this research. This includes the systematic calibration, simulation, modelling, and scaling of conventional MOSFET devices. The process and device simulation techniques, which have been used

to calibrate and simulate the Toshiba 35 nm MOSFET, are explained in detail. The different models employed and their specific parameters are explained in detail.

The results and discussions on the calibration, scaling and simulation of the scaled devices are presented in chapter four. The device structures of both the n and p -channel real 35 nm MOSFETs are described. This is followed by a description of the structure and the comparative characterisation of the scaled devices with 25, 18, 13, and 9 nm channel lengths. Additionally, the carrier mobility in the 35 nm and scaled transistors is studied in conjunction with the electric fields and doping concentrations at the Si/SiO₂ interface.

One of the major limiting factors of conventional MOSFET scaling are the intrinsic parameter fluctuations resulting from the position and number of random discrete dopants in the active region of the transistor or from gate line edge roughness. The statistical investigation of the impact of both the discrete random dopants and gate line edge roughness on the performances of 35 nm and scaled (25, 18, 13, and 9 nm) MOSFETs are presented in chapter 5.

There is a general consensus that the scaling of conventional MOSFET is nearing its end. It is important that the semiconductor industry looks for new technologies and device architectures. One of the promising emerging devices is the Ultra Thin Body (UTB) SOI FET. In the course of this research we have investigated the electrical properties of UTB SOI FETs under different scaling scenarios, which correspond to the 32 nm and 22 nm technology nodes. Chapter 6 presents the results of the scaling of these Ultra-Thin Body SOI FETs.

Finally, Chapter 7 of this thesis presents a summary of the main findings of the PhD research. In addition to the concluding remarks some suggestions for future research work are also outlined, highlighting some of the important research areas which may offer further insight into the simulation understanding and design of decananometre MOSFETs.

1.5 Chapter references

- [1.1] M. Eckert, and H. Schubert, Crystals, Electrons, Transistors: From Scholar's Study to Industrial Research, American Institute of Physics, 1990.
- [1.2] E. Braun, and S. Macdonald Revolution in Miniature: The History and Impact of Semiconductor Electronics, 2nd Edition, Cambridge University Press, 1982
- [1.3] J. S. Kilby, "Invention of the Integrated Circuit", *IEEE Trans. Electron Devices*, Vol. 23(7), pp. 648-654, 1976
- [1.4] R. C. Platzek, and J. S. Kilby, "Minuteman Integrated Circuit-A Study in Combined Operations", *Proceedings of the IEEE*, Vol. , pp. 1669,
- [1.5] J. S. Kilby, "The Electrical Century: The Integrated Circuit Early History", *Proceedings of the IEEE*, Vol. 88 (1), 2000
- [1.6] R.H. Dennard, F. H. Gaensslen, L. Khun and, H. N. Yu, "Design of micron switching devices", *Presented at IEDM*, Washington D.C., December 1972
- [1.7] E. J. Nowak, "Maintaining the benefits of CMOS scaling when scaling bogs down", *IBM J. Res. & Dev.* Vol. 46, NO. 2/3 2002
- [1.8] G. E. Moore, "Cramming More Components onto Integrated Circuit", *Electronics*, Vol. 38, No. 38, April 19 1965
- [1.9] M. C. Verlinde and Steven M. King, "Seeing Beyond Moore's Law", *Semiconductor International*, Vol. 25, No. 8, pp. 50, July 2002
- [1.10] T. Skotnicki, J. A. Hutchby, Tsu-Jae King, H. -S. P. Wong, and F. Boeuf, "The End of CMOS Scaling: Toward the Introduction of New Materials and Structural Changes to Improve MOSFET Performance", *IEEE Circuits & Device Magazine*, pp. 16-26 January/February 2005
- [1.11] A. Asenov, "Random Dopant Induced Threshold Voltage Lowering Fluctuations in sub-0.1 μm MOSFETs: A 3-D "Atomistic" Simulation Study" *IEEE Trans., Electron Devices*, Vol. 45, No. 45 pp. 2505-2513, December 1998
- [1.12] Semiconductor Industry Association, *International Technology Roadmap for Semiconductor Industry: 2003 edition*, Austin, Texas: 2003

2

The scaling of MOSFETs, Moore's law, and ITRS

For the last three decades, the microelectronic industry has benefited enormously from the MOSFET miniaturization. The shrinking of transistors to dimensions below 100 nm enables hundreds of millions transistors to be placed on a single chip. The increased functionality and reduced cost of large variety of integrated circuits and systems has brought its own benefit to the end users and above all the semiconductor industry. A low cost of manufacturing, increased speed of data transfer, computer processing power and the ability to accomplish multiple tasks simultaneously are some of the major advantages gained as a result of transistor scaling.

This chapter has four main sections. The first section deals with the Moore's law and its impact on the overall development of semiconductor technology and on MOSFET scaling in particular. Then the contributions made by the International Technology Roadmap for Semiconductors (ITRS) to the advancement of microelectronics technology from the MOSFET point of view are briefly discussed. At the same time the influence of the ITRS on the current priorities and directions related to the scaling of transistors will be discussed.

Section two describes the two basic forms of scaling considered by industry and research communities. Some of the fundamental limitations that that will eventually limit the scaling of conventional MOSFETs are examined in section three. The chapter ends with a summary presented in section four.

2.1 *The impact of Moore's law and ITRS on device scaling*

Moore's 'law' and the ITRS have been complimenting each other since the first edition of the roadmap in the early 90's. The former has been cast as a law from engineering observation made by G. Moore in the mid sixties [2:1]. It was initially a forecast on the number of transistors that can be integrated into a microchip for the next ten years (1965-1975), but the trend remained almost unchanged over the next three decades. The ITRS on the other hand is a comprehensive guide that enables the semiconductor industry to transform this observation into reality. At this stage, however, one has to be careful when interpreting "Moore's law", as a physical or mathematical law. Despite the efforts made by Meindle [2.2] to formulate the "compact mathematical formulation of the Moore's law" ($N = F^{-2} D^2 P_E$ where N is the number of transistors per chip, F is the minimum feature size, D is the chip area, and P_E is transistor packaging efficiency measured per minimum feature area), it remains simply an empirical observation on the rate of growth of semiconductor technology [2.3] originating from the forecast depicted inset to figure 2:1. Therefore, in order to clarify its role on the growth of semiconductor industry, in the following sub sections Moore's law is discussed briefly together with the ITRS mainly from the MOSFET scaling point of view.

2.1.1 Moore's Law

Back in time when Gordon Moore published his article, "Cramming more components onto integrated circuits" in 1965 [2:1], he was probably not aware of its impact on the remarkable progress of semiconductor technology in the years to come. In this publication he made an observation that it will be possible to integrate 6.5×10^4 components into a single chip by 1975, provided that the number of active transistor per chip doubled roughly every year. As illustrated in figure 2:1 the advances of the semiconductor technology have been able to follow this predicted trend.

When G. Moore made his prediction, the number of transistors in a single chip was roughly 32 and today there are approximately half a billion transistors integrated on a single microprocessor (figure 2:1). This phenomenal growth has demonstrated how

visionary his prediction was, and how vital has it been to the technology enabling the shrinking of individual transistors. The scaling of MOSFETs, which are the key components in digital technology, has revolutionized the semiconductor industry and has also enabled the realization of the immensely complex devices and systems we rely on at present.

Although the “Moore’s law” has been interpreted differently at the different stages of the semiconductor technology industry’s development, the formulation that has been accepted as a general consensus states that: “the number of components per chip doubles every 18 months” [2.4]. Note that the original assumption made by Moore, according to the inset in figure 2.1, was that the number of components per chip will be doubled every 12 months. Indeed the originally stated rate of development was maintained in the seventies, as shown by Moore himself in 1975 [2.5], and continued to the early eighties. The present 18 months period of doubling of the chip components is a modification in line with the past and present (2003) ITRS editions and the real state of the industry.

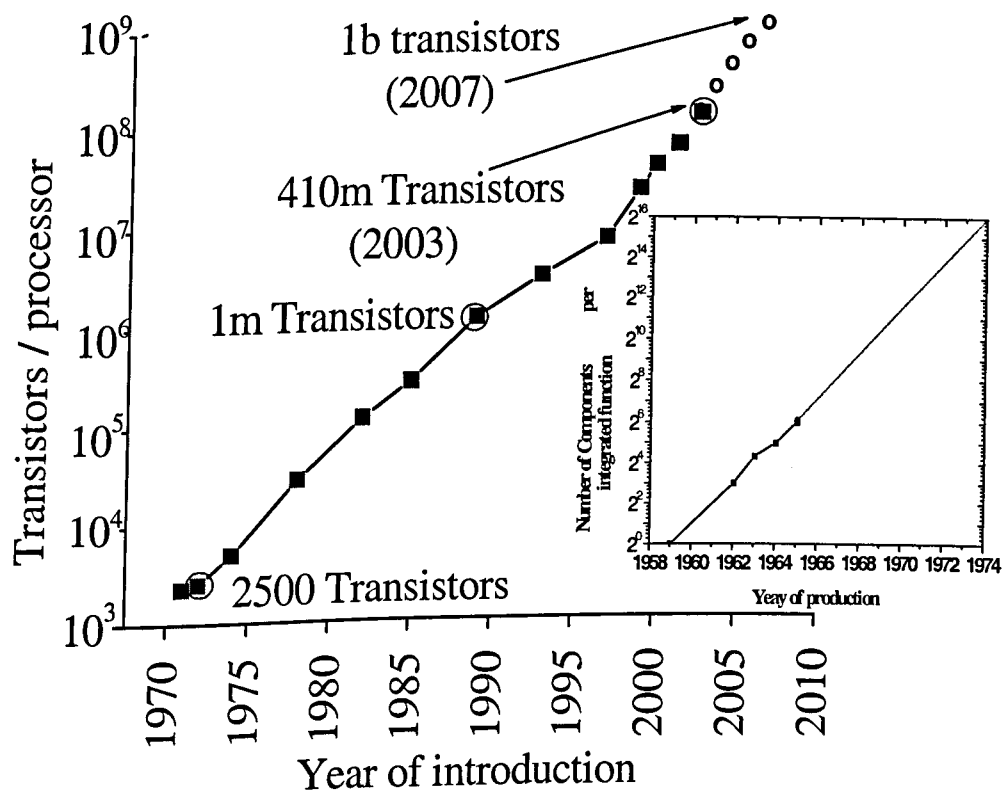


Figure 2.1 Visualization of Moore’s Law: The number of transistors integrated in a commercially available processor and the outlook towards a billion transistors in a single processor due in year 2007 (Intel). The inset graph: Projection made by Moore on his original paper on the number of components per integrated device [2.1]

2.1.2 Implication of Moore's law

Moore's law has had various implications on the microelectronics manufacturing industry and user applications in general over the last 30 years. As a result, increasing functionality [2.6], cost per function reduction, and better performance, have all been achieved for every new generation of integrated circuits.

According to the ITRS, the functionality is defined as the number of bits in a DRAM chip or the number of logic transistors in a microprocessor unit. With the integration of more individual components in a single chip the functionality per chip increases (figure 2:2) together with the increase in the density of functions (functions/area). The increase in functionality minimizes the delay of data flow that occurs due to the isolation of individual functions on separately integrated systems [2.7]. More functionality also means an increase in overall physical density of useable transistors per total chip area. Figure 2:2 shows according to the ITRS, that in both the near-term (2003-2009) and the long-term (20010-2018) functionality will be increasing by roughly 100% in every technology nodes.

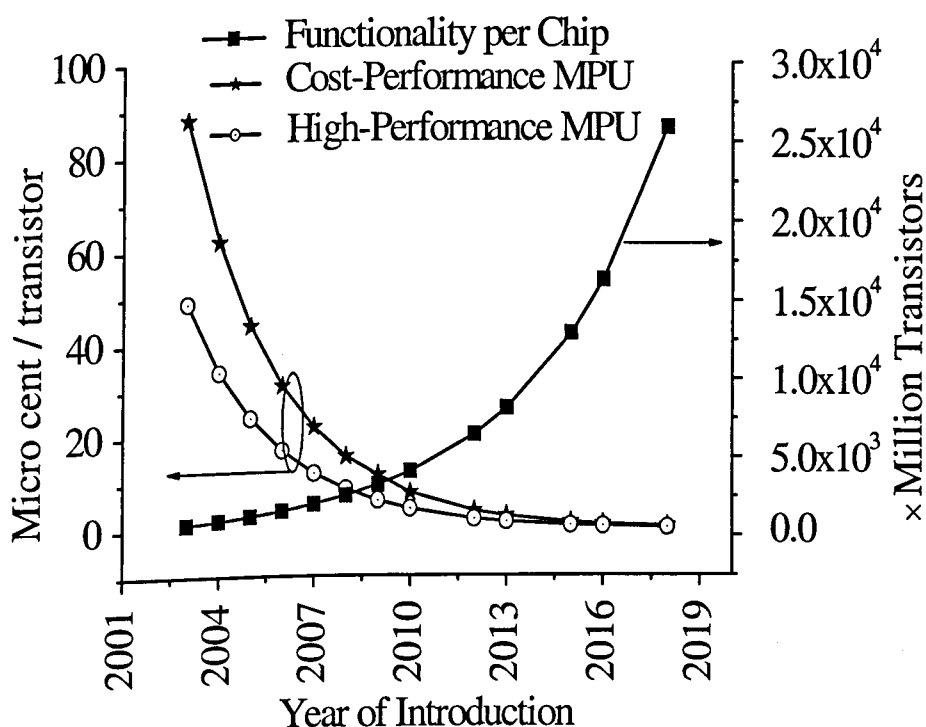


Figure 2:2 Cost – performance of Microprocessor Unit (MPU), cost of high speed performance MPU and functionality (functionality is often associated with the number of bits or unit devices in MPU) against the year of introduction of technology nodes: data ITRS 2004 update.

The second important feature associated with the Moore's law is cost. It is a general rule that the goal of every manufacturing community is maximizing the profit while minimizing the cost of production. The electronics industry is not unique in this. In fact the primary implication of the Moore's law is the reduction of manufacturing cost per function and at the same time to increase the functionality per chip. As it can be seen from figure 2:2 the reduction in cost-per-function according to the latest ITRS edition, is roughly 50% in about two years.

The third important implication of Moore's law is the performance factor. Performance in general can be measured, for example, by the speed of typical microprocessors. Figure 2:3 shows the increasing speed and density of present and future generations of technology nodes. The off-chip frequency is the maximum input and output signal frequency to board peripheral buses of high performance devices [2.4]. The off-chip frequency is increasing faster than the on-chip local frequency near the end of the current edition of the ITRS.

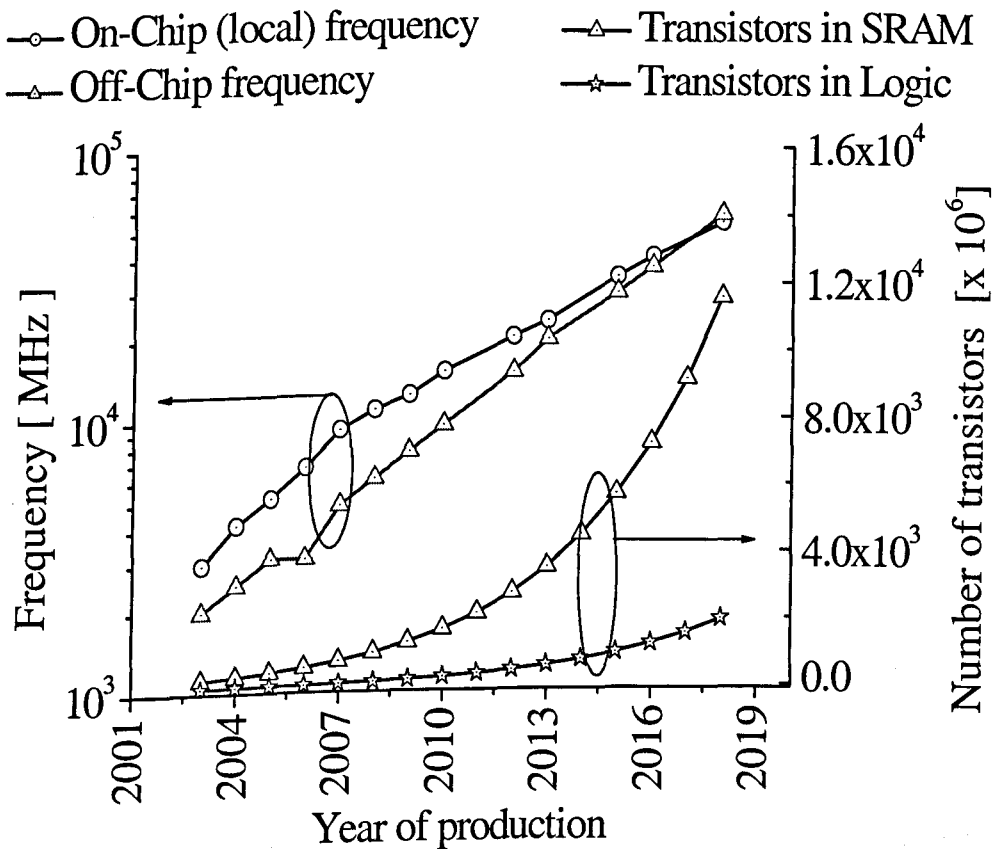


Figure 2:3 The technology trends of on-chip local clock, off-chip frequency, chip density in SRAM and Logic gates (transistors per cm^2). ITRS 2004 Edition

In general the frequency is related to the speed of switching of the individual and simple logic components determined by CMOS transistors which is inversely proportional to the delay time that takes to propagate signal through the inverter. The inverter delay time formulated as [2.8]:

$$\tau_{int} = R_{sw}(C_{in} + C_{out}) \quad (2.1)$$

Where τ_{int} is the inverter delay time, R_{sw} is the switching resistance, C_{in} input capacitance, C_{out} the output capacitance, and in equation 2.2, C_{gate} is the gate capacitance. This inverter delay time can be used as approximation of the CMOS delay time which is calculated empirically as: [2.9]

$$\tau = C_{gate} \frac{V_{DD}}{I_{dsat}} \quad (2.2)$$

According to the generalised scaling [2.10, 2.11], τ is inversely proportional to the scaling factor, which allows faster circuit operations. Figure 2:2 also depicts the increase in the density of transistors in SRAM and Logic circuits. Density is inversely proportional to the total chip area ($1/A$). Therefore, density will increase by κ^2 as a result of scaling, where $\kappa \approx \sqrt{2}$ is the scaling constant (see section 2.1.4).

2.1.3 The International Technology Roadmap for Semiconductors

The technology roadmap is an ambitious document widely used as a guiding reference for advanced semiconductor device research and manufacturing purposes. The latest edition (2003)[†] of international technology roadmap for semiconductors (ITRS), updated in 2004, sets main objectives and targets to 2018. Based on research from the semiconductor industry and academia, the latest edition of the ITRS outlines the requirements and identifies the challenges which allow Moore's law to be maintained over the next 15 years. In addition to the challenges, it also outlines the possible solutions to some of the problems that the industry may face and highlights the specific areas that need urgent research.

Overall, the roadmap has three major contributions. The first is to identify the needs and requirements to be met by technology solutions currently under development. The second is to recognize the existence of interim solutions for the medium term challenges and problems and their limitations at the present time. The third important

[†] The latest ITRS edition of 2005 has been published after the completion of this thesis

contribution of ITRS is to identify the areas where there are “no known manufacturing solutions” customarily labelled as the “Red Brick Wall” - to induce the industry to concentrate on them strategically and focus research efforts in these areas.

The ITRS is a comprehensive document with more than 600 pages, including the executive summary. It covers fifteen categories related to semiconductor industry and to basic research and development areas. One of the important sections expanded significantly in the latest edition ITRS is on the emerging research devices. It was organized with the aim of finding and building successful new device structures that can replace conventional MOSFETs. Although some of the listed structures are more of research type, the device structures such as fully depleted silicon on insulator (FD SOI) and the multiple gates architectures including the double gate MOSFETs and FinFETs are the promising candidates to replace mainstream device structures.

Year of Production	2003	2004	2005	2006	2007	2008	2009
Technology Node	hp90			hp65			
DRAM $\frac{1}{2}$ Pitch (nm)	100	90	80	70	65	57	50
MPU/ASIC M1 $\frac{1}{2}$ Pitch (nm)	120	107	95	85	75	67	60
MPU/ASIC Poli Si $\frac{1}{2}$ Pitch (nm)	100	90	80	70	65	57	50
MPU Printed L_g (nm)	65	53	45	40	35	32	28
MPU Physical L_g (nm)	45	37	32	28	25	22	20
Equivalent t_{ox} (nm)	1.3	1.2	1.1	1	0.9	0.8	0.8
V_{dd} (HP)	1.2	1.2	1.1	1.1	1.1	1	1
Off current, I_{off} [$\mu A/\mu m$]	0.03	0.05	0.05	0.05	0.07	0.07	0.07
Drive current, I_{on} [$\mu A/\mu m$]	980	1110	1090	1170	1510	1530	1590
HP NMOS intrinsic delay τ [ps]	1.2	0.95	0.86	0.75	0.64	0.54	0.48
Relative intrinsic speed, $1/\tau$	1	1.26	1.39	1.6	1.86	2.2	2.49
Logic gate delay [ps]	30.24	23.94	21.72	18.92	16.23	13.72	12.13
DRAM cell size [μm^2]	0.082	0.065	0.048	0.036	0.028	0.019	0.015
S/D extension x_j [nm]	24.8	20.4	17.6	15.4	13.8	8.8	8.0

Table 2:1 The near term years (2003-2009) of selected overall roadmap technology characteristics that are required to continue the present scaling trends of conventional MOSFETs. Half pitch 90 and 65nm technology nodes are marked as hp90 and hp65 respectively. (ITRS 2003 edition)

However, not all of these categories and data are relevant to this research. Therefore, in this sub-section we only concentrate on the high performance devices, which are in the heart of this work. The summarised data of device dimensions and electrical parameters for high performance devices depicted in tables 2:1 (near-term years) and 2:2 (long-term years) have been adopted as a guide for the scaling of the 35 nm MOSFET described in chapter 4.

The carefully calibrated 35 nm gate length MOSFETs manufactured by Toshiba [2.12]) were used as a basis for further scaling to gate lengths of 25, 18, 13, and 9 nm transistors. The overall calibration and scaling methodology and results are presented in chapter 3 and 4 respectively. The dimensions of the 35 nm MOSFET physical gate length used for this work are not characteristics of particular node on the ITRS roadmap. It's performance, $I_{on} = 676\mu\text{A}/\mu\text{m}$, $I_{off} = 100\text{nA}$ at $V_{dd} = 850\text{mV}$ and design parameters, $t_{ox} = 1.2\text{ nm}$ $x_j = 20\text{ nm}$ are close to the 37 nm high performance device required for the 90 nm node and 80 nm technology generations.

Year of Production	2010	2012	2013	2015	2016	2018
Technology Node	hp45		hp32		hp22	
DRAM $\frac{1}{2}$ Pitch (nm)	45	35	32	25	22	18
MPU/ASIC M1 $\frac{1}{2}$ Pitch (nm)	54	42	38	30	27	21
MPU/ASIC Poli Si $\frac{1}{2}$ Pitch (nm)	45	35	32	25	22	18
MPU Printed L_g (nm)	25	20	18	14	13	10
MPU Physical L_g (nm)	18	14	13	10	9	7
Equivalent t_{ox} (nm)	0.7	0.7	0.6	0.6	0.5	0.5
V_{dd} (HP) (V)	1	0.9	0.9	0.8	0.8	0.5
Off current, I_{off} [$\mu\text{A}/\mu\text{m}$]	0.1	0.1	0.3	0.3	0.5	0.5
Drive Current, I_{on} [$\mu\text{A}/\mu\text{m}$]	1900	1790	2050	2110	2400	2190
HP NMOS intrinsic delay τ [ps]	0.39	0.3	0.26	0.18	0.15	0.11
Relative intrinsic speed, $1/\tau$	3.06	4.05	4.64	6.8	8.08	10.77
Logic gate delay [ps]	9.88	7.47	6.55	4.45	3.74	2.81
DRAM cell size [μm^2]	0.1222	0.0077	0.0061	0.0038	0.0025	0.0016
S/D extension depth x_j [nm] [‡]	7.2	11.2	10.4	8.0	7.2	5.1

Table 2:2 The long - term years (2010-2018)

[‡] The extension depth (x_j) is calculated with the assumption of introducing new device structures beyond year 2007, like fully depleted SOI and multi gate device structures. (ITRS 2003 edition)

Although the electronics industry prefers to continue as long as possible with the scaling of conventional MOSFETs, there is “Red Brick Wall” to this process unless there is a major technological breakthrough. High channel doping, which degrades the device performance, and ultra thin gate oxides, which introduce unacceptable gate leakage, are likely to prompt a replacement to conventional MOSFETs somewhere beyond the 65nm technology node. Among the replacement candidates are, for example, ultra-thin body SOI and multiple gate devices complimented by the introduction of strained silicon in the channel region to enhance the carrier mobility, and high permittivity materials in the gate stack in order to suppress gate leakage. Some of the critical scaling limitation factors will be examined more closely in the next sections of this chapter.

2.1.4 The scaling factors and technology trends

The scaling factor of $\kappa \approx \sqrt{2}$, related to a 70% size reduction of the major technology nodes every two years, has been adapted for the linear scaling of device dimensions in this work. The other scaling constant, α for the electric field and potential used in the generalised scaling scenario [2:10] is not specified on the road map. However, it can be calculated from the supply voltages (V_{dd}), which are specified in the roadmap for corresponding feature sizes and the linear scaling factor κ as:

$$V' = \frac{\alpha}{\kappa} V \Rightarrow \alpha = \kappa \frac{V'}{V} \quad (2.3)$$

V' is the new supply voltage given in the technology roadmap and V is the supply voltage of the previous generation. It should be noted that in some papers [2.13], the linear scaling factor has been decomposed to separate dimensional scaling parameters in the so called “selective scaling case”, which introduces different values for vertical, horizontal, and lateral dimensions multipliers. However, in this work, the generalized scaling rule has been adopted as a principal guiding rule for device scaling. A review of the different scaling approaches is presented in the next section.

Unlike the previous editions of ITRS, no prediction of the technology acceleration has been made in its latest edition (ITRS'03). Also, as illustrated in figure 2.4, in the last ITRS edition, the technology generations are predicted to shift from the present two-year cycle to a three-year cycle trend around 2007. The technology node

continued to be defined as 70% dimension reduction per node or approximately 50% reduction per two nodes. The “technology-node-cycle” is the period of time in which a new technology node is introduced.

In addition to the scaling of the gate length, the oxide thickness is another critical parameter, which has been aggressively scaled down in order to achieve a sufficient drive current and to control short channel effect. The later can be achieved by maintaining the electrostatic control of the channel potential by the gate.

Figure 2:4 shows the technology half pitch (hp) and gate length trends adopted in the ITRS’03 edition. Beyond year 2007 the two year cycle delays by another year and is expected to be three years until the end the present roadmap projection time-line and probably beyond. The physical gate length is conventionally adapted as minimum feature size regarding the individual devices.

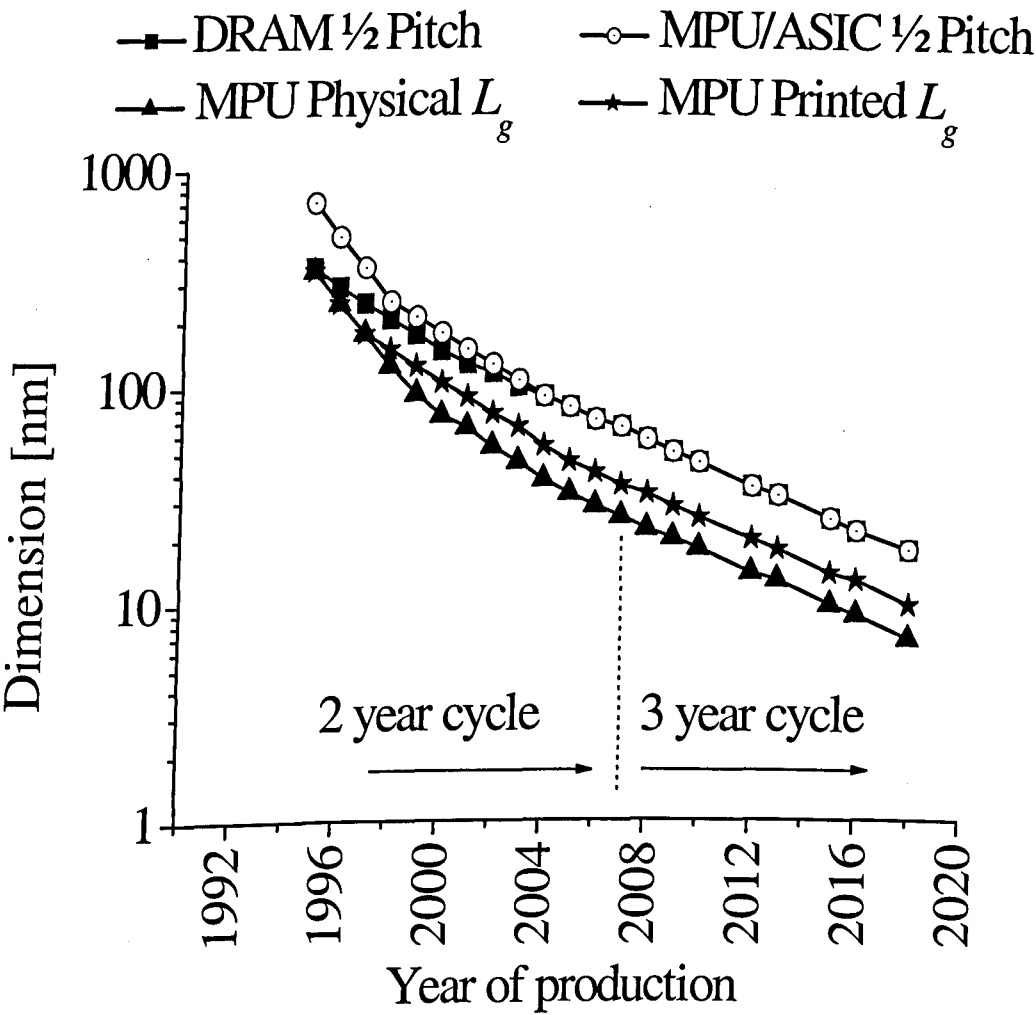


Figure 2:4 Technology half-pitch and gate length trends.

2.2 The scaling rules for conventional MOSFETs

In the preceding sections the technology roadmap and the Moore's law have been discussed in order to examine their role in pursuing transistor scaling, and above all in highlighting scaling's unprecedented contributions to the enormous advance of semiconductor technology. Without the extraordinary miniaturization of transistors, it would be impossible to produce higher volumes of faster devices operating at lower power. Nobody in the semiconductor industry disputes this state of affairs. This section further introduces the theory and practice of the scaling process. It begins by reviewing some of the classic papers on the constant field and generalised device scaling rules, followed by detailed analysis of advantages and shortcomings of both rules.

2.2.1 Constant field scaling

Dennard *et al.* presented their pioneering research work on the scaling of MOSFET devices at the International Electron Device Meeting (IEDM) 1972 [2.14] and published a comprehensive paper on the scaling of MOS transistors in 1974 [2.15], from which the "constant field scaling" theory has emerged. The basic principle which they employ is that in order to increase the performance of a MOSFET we must reduce linearly the size of the transistor, together with the supply voltage, and increase the doping concentration in a way which keeps the electric field in the device constant - hence the name "constant field scaling" (figure 2.5).

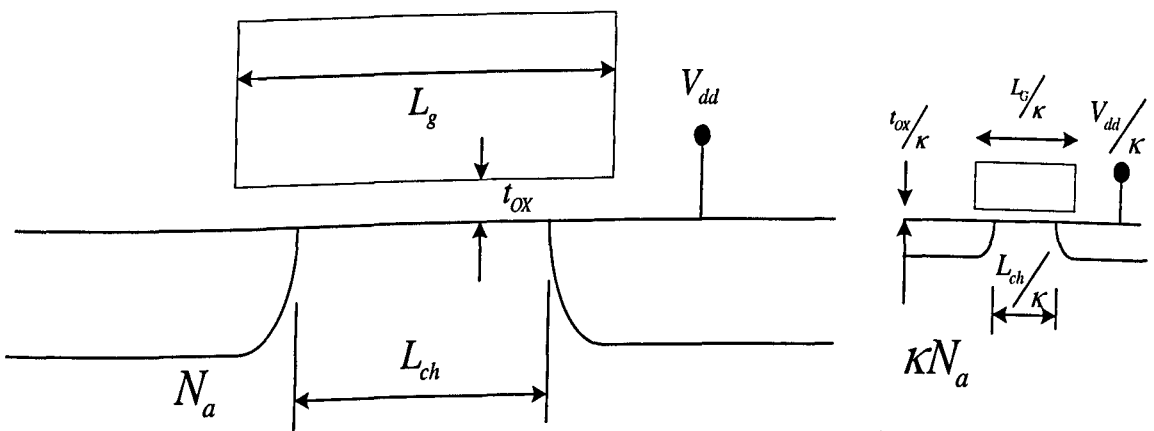


Figure 2: 5 Illustration of MOSFET miniaturisation. The sketch on the right hand is the scaled device according to the constant field rule. (Reference [2.15])

The scaling process is performed by a linear transformation of three design parameters (voltage, doping concentration, and physical dimensions) of a particular generation of transistor by the same scaling factor, κ . The scaled down device will have a reduced voltage (V_{dd} / κ), vertical (t_{ox} / κ and x_j / κ) and horizontal (L / κ) dimensions, and an increased doping concentration (κN_a) as depicted in figure 2.5. Despite the change in those parameters, the intensity of the electric field ($v/L = v'/L'$) remains virtually unchanged since both the dimension and the supply voltage scale by the same ratio.

Table 2.3 summarises the changes in device dimensions and circuit parameters as a result of both the constant field and the generalised scaling rules. Only the most important design and operational parameters are included in the table. The other quantities which are not given in table 2.3 can be deduced using the listed design parameters.

Scaled Parameters	Constant field Scaling	Generalised scaling
t_{ox}, L, W, X_j, W_d	$1/\kappa$	$1/\kappa$
N_a, N_d (ions/cm ³)	κ	$\alpha\kappa$
Power supply: (V_{dd})	$1/\kappa$	α/κ
Electric field in device: (E)	1	α
Capacitance: (C)	$1/\kappa$	$1/\kappa$
Inversion charge density (Q)	1	α
Circuit delay time: $\tau \sim C/I$	$1/\kappa$	$1/\kappa$
Power dissipation: (P)	$1/\kappa^2$	α^2/κ^2
Power density ($\sim P/A$)	1	α^2
Circuit density	κ^2	κ^2
Chip Area (A)	$1/\kappa^2$	$1/\kappa^2$
Current, Drift: (I)	$1/\kappa$	$1/\kappa$

Table 2:3 Summary of the constant field scaling and the generalised scaling rules

The other issue addressed in [2.15] was the application of ion implantation in the fabrication process of the scaled device. This is an important process step which allows us to more accurately place the dopants in the shallower source drain junctions and channel of the scaled device. In addition to its advantages, the shortcomings of the scaling process have also been addressed in [2.15]. For example, carrier mobility degradation as a result of high doping in the channel and the short channel effect (decreasing of V_T) are some of the main drawbacks. The adverse effect of doping concentration on carrier mobility can be observed from the empirical formula given by equations (2.4 & 2.5) [2.16].

$$\mu_e = 88T_n^{-0.57} + \frac{7.4 \times 10^8 T_r^{-2.23}}{1 + \left[\frac{N_d}{1.26 \times 10^{17} T_n^{2.4}} \right] 0.88 T_n^{-0.146}} \quad (2.4)$$

$$\mu_p = 54.3T_p^{-0.57} + \frac{1.36 \times 10^8 T_r^{-2.23}}{1 + \left[\frac{N_a}{2.35 \times 10^{17} T_p^{2.4}} \right] 0.88 T_p^{-0.146}} \quad (2.5)$$

Where, T_n and T_p are the electron and hole temperatures of interest respectively and T_r is the room temperature. μ_n and μ_p are the electron and hole mobility respectively. N_a and N_d are acceptor and donor concentration. Since carrier mobility is inversely proportional to the channel doping as shown in the equations, increasing channel doping reduces the mobility and the device performance. Regardless of implementation of different channel engineering techniques, the highly doped channel imposes carrier transport problems in aggressively scaled MOSFETs.

2.2.2 Generalised scaling rule

As device dimensions enter into the sub-micron dimensions, two-dimensional effects (short channel effect-SC and drain induced barrier lowering-DIBL) become increasingly important. The gradual field approximation becomes invalid and the field changes significantly even if the constant field scaling scenario is applied. The field also increases due to a much slower reduction in the supply voltage in real circuits compared to the requirements of constant field scaling. This challenge to constant field scaling has

been addressed by Brews *et al.* [2:11], and Baccarani *et al.* [2:10], who have introduced a more generalised scaling theory.

Brews *et al.* mainly concentrated on the minimum channel length for which the subthreshold characteristics of the long channel device can be maintained in the scaled devices. For this purpose they suggested an empirical formula given by:

$$L_{\min} = A \left[x_j t_{ox} (W_s + W_d)^2 \right]^{\frac{1}{3}} \quad (2:6)$$

Where L_{\min} is the minimum channel length, W_s and W_d are the depletion widths in the source and drain regions respectively, x_j is the junction depth, A and t_{ox} are proportionality constant and oxide thickness respectively.

The main advantage of this approach over the constant field scaling, according to [2.11], is that the parameters do not all have to be scaled by one factor. But there is a drawback associated with the way in which the minimum channel length is determined. It was suggested that the channel length could be reduced until a 10% increase in drain current is obtained.

However, the tolerance to the short channel effects may not only depend on a predetermined drain current value, but also depend on circuit applications [2.17]. In addition to this, the scaling of a MOSFET includes to at least five major design parameters (L_g , t_{ox} , V_{dd} , N_a , x_j) [2.8] [2.18], not just the three defined in equation (2.3). The typical electrical behaviour of device under the influence of short channel effect (SCE) and drain induced barrier lowering (DIBL) are highly dependant on all five parameters as shown in equation (2.7 and 2.8)[§] [2.19] [2.16].

$$SCE = 0.64 \frac{\epsilon_{si}}{\epsilon_{ox}} \left(1 + \frac{x_j^2}{L_{el}^2} \right) \frac{t_{ox}}{L_{el}} \frac{W_{dm}}{L_{el}} V_{bi} \quad (2.7)$$

$$DIBL = 0.80 \frac{\epsilon_{si}}{\epsilon_{ox}} \left(1 + \frac{x_j^2}{L_{el}^2} \right) \frac{t_{ox}}{L_{el}} \frac{W_{dm}}{L_{el}} V_{ds} \quad (2.8)$$

[§] For further reference on the derivations of equations 2.7 and 2.8, please look in [2.18]

V_{bi} and V_{ds} are the built in potential and the input drain Voltage. The effective channel length is defined as $L_{el} = L_g - \Delta L$, where L_g is the physical gate length and ΔL is the sub-diffusion length. It is also clear that from the empirical formulas of CMOS design rules, which require $x/L \approx 0.33$, $t_{ox}/L \approx 1/30$, $w_d/L \approx 0.33$, $v_T/V_{dd} \approx 0.20$, that all the five parameters influence the electrostatic integrity of the scaled devices, which determines both SCE and DIBL.

On the other hand, by identifying a significant difference in the two dimensional pattern of the electric field in the active region of the original and the scaled device, Baccarani *et al.* have suggested that the supply voltage and the doping concentration should be scaled with different scaling factor. To facilitate this a new scaling concept an additional scaling constant, α , has been suggested. The effect of α can be demonstrated by examining the Poisson equation within the depletion region, which is explained in [2:8] and given by the equation:

$$\frac{\partial^2(\alpha\psi/\kappa)}{\partial(x/\kappa)^2} + \frac{\partial^2(\alpha\psi/\kappa)}{\partial(y/\kappa)^2} = \frac{qN'_a}{\epsilon_{si}} \quad (2.8)$$

Where $N'_a = \alpha\kappa N_a$, is the channel doping concentration in the scaled device. Equation (2.8) is based on the assumption that the potential will be scaled by α/κ and the electric field by just α (where $\alpha \geq 1$). It is, however, important to note that in [2.17] the scaled potential is given as $\psi' = \psi/\kappa$ which is different than the one shown in equation (2.3). Moreover scaling the potential by the same factor as the dimensions leads to the constant field scaling theory. The main reason for adopting in this work the generalised scaling with α determined by equation (2.3) is the fact that the supply voltage can not be scaled as fast as the device dimensions due to the non-scaling property of the threshold voltage and the subthreshold slope [2.10].

The main problem with the generalized scaling rule, particularly in deep sub-100 nm scaled devices is an increase on power density ($P/A \Rightarrow \alpha^2$). The scaling of total area scales as $1/\kappa^2$ while the power dissipation per circuit scales as α^2/κ^2 , i.e., the size of the area scales down faster than the power dissipation. This difficulty is considered to be one of the major scaling limitation factors [2.20]. The scaling limits are discussed in detail in later sections of this chapter.

2.2.3 Evolution of CMOS design

Although the history of semiconductor devices goes back to the 1920s, this thesis only concentrates on the period after 1960s where the transformation of metal oxide semiconductor field effect transistors (MOSFET) from research devices to the major building block of commercial integrated circuits takes place. For further reading, a comprehensive historical review is presented in [2.21].

The first working CMOS circuit was developed circa 1964 at RCA [2.22] and was integrated in to logic gates in late 1960’s [2.23]. The fabrication of both n-MOS and p-MOS transistors on the same wafer was an important stage in revolutionizing the integrated circuit. It has the advantages over using single *n* and *p*-MOSFET devices discussed next.

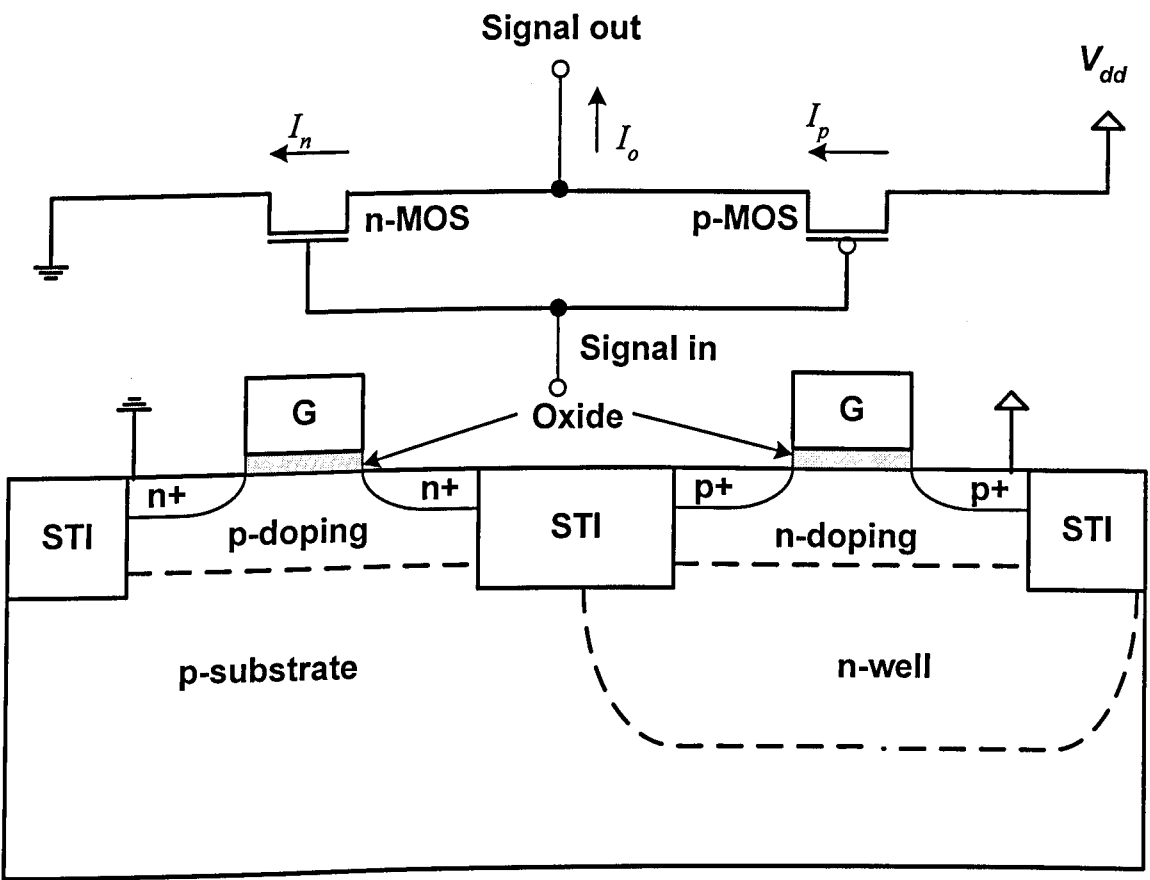


Figure 2:6 Cross-sections of fabricated CMOS device which can be used as inverter circuit. STI stands for shallow trench isolation and the currents I_n and I_p are drain currents of n and p-type devices and I_o is an output current of an inverter.

CMOS circuits offer high switching speed, high density of integration, and very low static power dissipation. These advantages favour the miniaturisation of MOSFETs and subsequent realization of high density integrated circuits with ever increasing speed. One of the simplest building blocks for CMOS logic gates is the inverter illustrated in figure 2:6 which is realized on a single wafer using single, double or triple well technology.

The next important milestone in the evolution of MOSFET design is the introduction of the self-aligned polycrystalline-silicon (poly) gate in the early 80's. The self-alignment of the source and drain to the gate reduces stray capacitance which improves the signal propagation delay, $\tau \sim C/V$ and overall circuit performance. Moreover polycrystalline-silicon as a material is stable and completely compatible with silicon technology [2.24]. The gate material work function must be suitable in order for the device to have an acceptable subthreshold voltage. Polysilicon has properties which match all these requirements.

As MOSFET channel lengths approach sub-micron dimensions the high electric field in the channel start to affect the device reliability and the introduction of lightly doped drain (LDD) MOSFETs in of late 70's are required [2.25] (see figure 2:7A). The lightly doped n^- region in the neighbourhood of the conventional n^+ source and drain areas soften the electric field and reduce hot carrier injection in the oxide [2.26] [2.27]. With the reduction of drain voltages the need for LDD subsides.

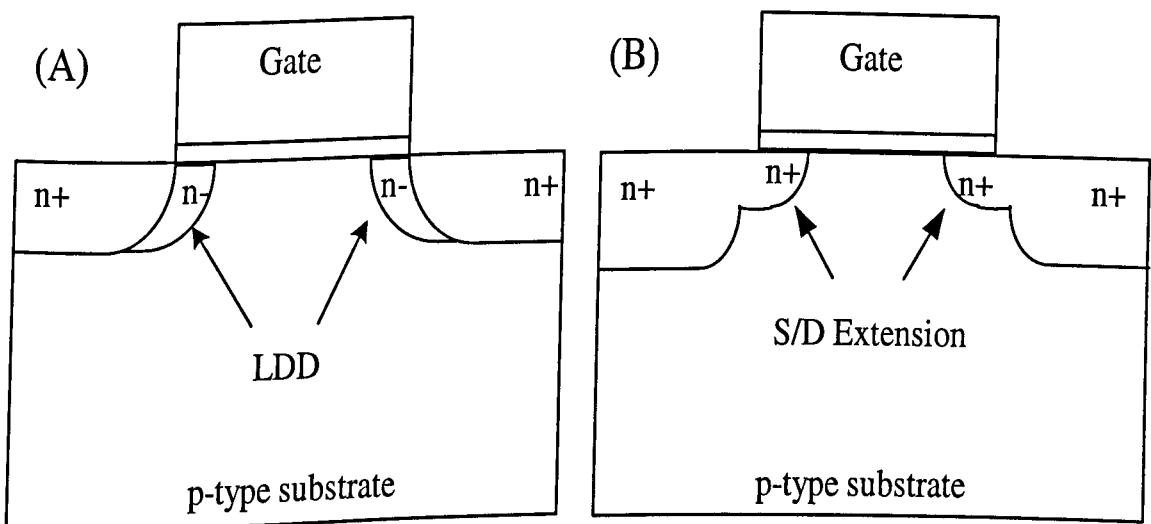


Figure 2:7 Cross section view of LDD n-MOSFET (A) and MOSFET with heavily doped source and drain extensions (B)

Later on, heavily doped but very shallow source and drain extensions (illustrated in figure 2:7B), were introduced to combat short channel effects without introducing problems associated with high series resistance [2.28].

One of the important technology stages during the process optimization and performance enhancement of the individual MOSFET is the introduction of Salicide (Self aligned) into the CMOS production process (see figure 2:8). The increase of the resistance of the poly-gate and the parasitic resistance of the shallow source and drain junction results in a poor performance of the scaled down MOSFETs. Self-aligned silicide technology has been first suggested by Crowder *et al* to be used as a shunting gate electrode on top of poly-gate [2.29].

In late 80's its application expanded to the source and drain electrodes to reduce the access and the contact resistance. Since the sheet resistance increases with the reduction of junction depth, it becomes important to use silicide to reduce the parasitic source and drain resistance in order to achieve the required drive current. TiSi_2 and CoSi_2 are the most widely used metal silicide in semiconductor industry to day, with the resistivity of 13-16 and 22-28 $\Omega\text{-cm}$ respectively. NiSi_2 is the next promising candidate.

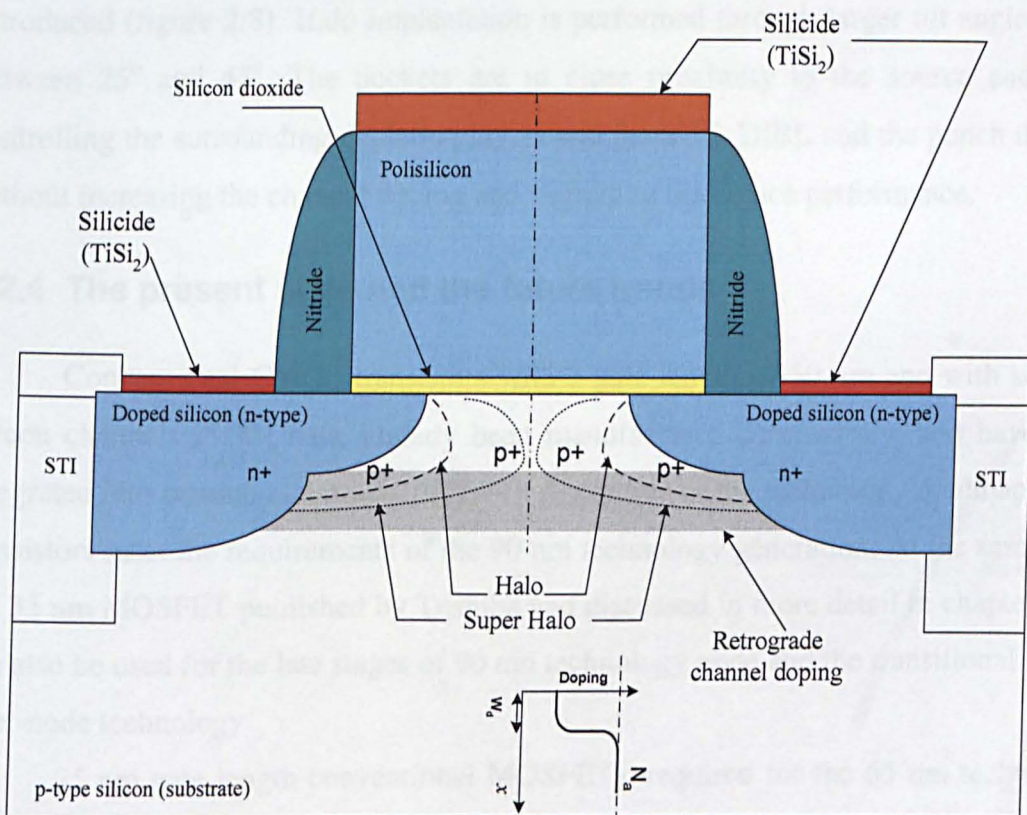


Figure 2:8 schematic illustrations showing self-aligned silicide in source/drain and on the poly-gate, halo (super-halo) doping and retrograde channel doping in a cross sectional view of the MOSFET. The inset depicts the ideal low-high (retrograde) doping profile.

Another key aspect in the evolution of CMOS design is channel engineering. Although various doping schemes have been implemented throughout the history of CMOS technology development, only the retrograde and the halo channel doping which are adopted in the deep submicron MOSFETs designs will be reviewed here.

According to the generalised scaling rule, the channel doping increases $\kappa\alpha$ times in every successive technology generation, in order to control short channel effects. However, this increase in channel doping severely degrades carrier mobility. The traditional channel doping scheme which uses the flat well profile to control V_T and combat short channel effects is not applicable in 100 nm channel length MOSFETs. To overcome problems associated with device performance and to provide relatively lower threshold voltage the super steep retrograde (SSR) channel doping scheme was proposed in early 90's in [2.30]. SSR doping profile increases current drive by enhancing carrier mobility and reducing V_T . [2.31].

For sub 0.1micron transistors the SSR alone is not sufficient to improve the performance (I_{on}) and to control the short channel effect at the same time. To reduce this problem the halo (pocket) doping and latter the supper halo doping [2.32] were introduced (figure 2:8). Halo implantation is performed through larger tilt angles often between 25° and 45° . The pockets are in close proximity to the source and drain controlling the surrounding depletion layers and the SCE, DIBL and the punch through without increasing the channel doping and degrading the device performance.

2.2.4 The present state and the future trends

Conventional CMOS transistors with a gate length of 50 nm and with strained silicon channels [5:33] have already been manufactured successfully, and have been integrated into commercial products [5:34]. According to the technology roadmap, these transistors meet the requirements of the 90 nm technology generation. At the same time the 35 nm MOSFET published by Toshiba and discussed in more detail in chapter four, can also be used for the late stages of 90 nm technology node and the transitional 80 nm inter-node technology.

25 nm gate length conventional MOSFETs, required for the 65 nm technology node, have also been reported in [2:35]. It has good subthreshold characteristics ($S = 118$ mV/decade, $I_{off} = 100$ nA/ μm) and a high drive current ($I_{on} = 840\mu\text{A}/\mu\text{m}$). Such 25 nm transistors are expected to deliver the performance required by the ITRS. However there

is still a need for '*total process optimization*' [2:35] in order to use the current fabrication technology. For example, better designed doping profiles in the channel, improved gate patterning (minimum LER), shallow junction formation to reduce source-drain sheet resistance, and optimization of the gate oxide (oxynitride) by controlling the amount of nitrogen, are some of the process-optimization steps that can be exploited. In addition to process optimization, the introduction of strained silicon in the channel to enhance carrier transport, and introduction of high- κ materials to replace silicon dioxide as an insulator will improve the chances of using conventional MOSFETs at the 65 nm technology node and possibly beyond.

Apart from the 25 nm gate length MOSFET, which gives a realistic hope of realizing the 65 nm technology node by 2007, there have also been research demonstrations of smaller gate length devices. For example, conventional MOSFETs with gate lengths 15 nm [2.36], and 16 nm [2.37], which are required for hp45 nm technology node; 14 nm [2.38] for hp32 nm and 6 nm [2.39], 8 nm [2.40] and 10 nm [2.41] required for the hp22 nm technology node, have been fabricated and reported, delivering promising device parameters.

There is, however, growing consensus among the industry and research communities alike, that in the 45 nm technology node and beyond, it will become necessary to replace the conventional MOSFETs with novel device architectures, silicon on insulator (SOI), multiple gate FETs and wide application technology boosters such as strained silicon for carrier transport enhancement, high- κ gate stack, metal gates, etc.

2.3 Factors limiting the scaling of conventional MOSFET

Some of the technology advances enabled the scaling of conventional MOSFETs to decananometre dimensions were discussed in the previous sections. Unfortunately, even with all these advances the scaling of conventional MOSFETs whilst which maintaining good performance becomes increasingly difficult over time. Although the electronics industry has benefited from continuous scaling over the last three decades or so, the present trends indicate that the scaling of the conventional MOSFETs is fast approaching the end of its useful life time.

The optimistic prediction of the 2003 edition of the ITRS that scaling will continue until 2018 and beyond, is challenged by some fundamental limitations. Quantum mechanical effects such as carrier confinement and tunnelling, the

randomness of discrete doping, and worries over the increasing power dissipation are some of the main factors that may force the industry to a paradigm shift in MOSFET architecture and process technology. In the next section, some of the fundamental limitations to scaling are examined

2.3.1 Quantum mechanical tunnelling

The three main quantum mechanical tunnelling phenomena, which affect the MOSFET scaling, are illustrated in figure 2:9. They include band-to-band tunnelling, gate tunnelling, and source to drain tunnelling. All three tunnelling process are discussed below.

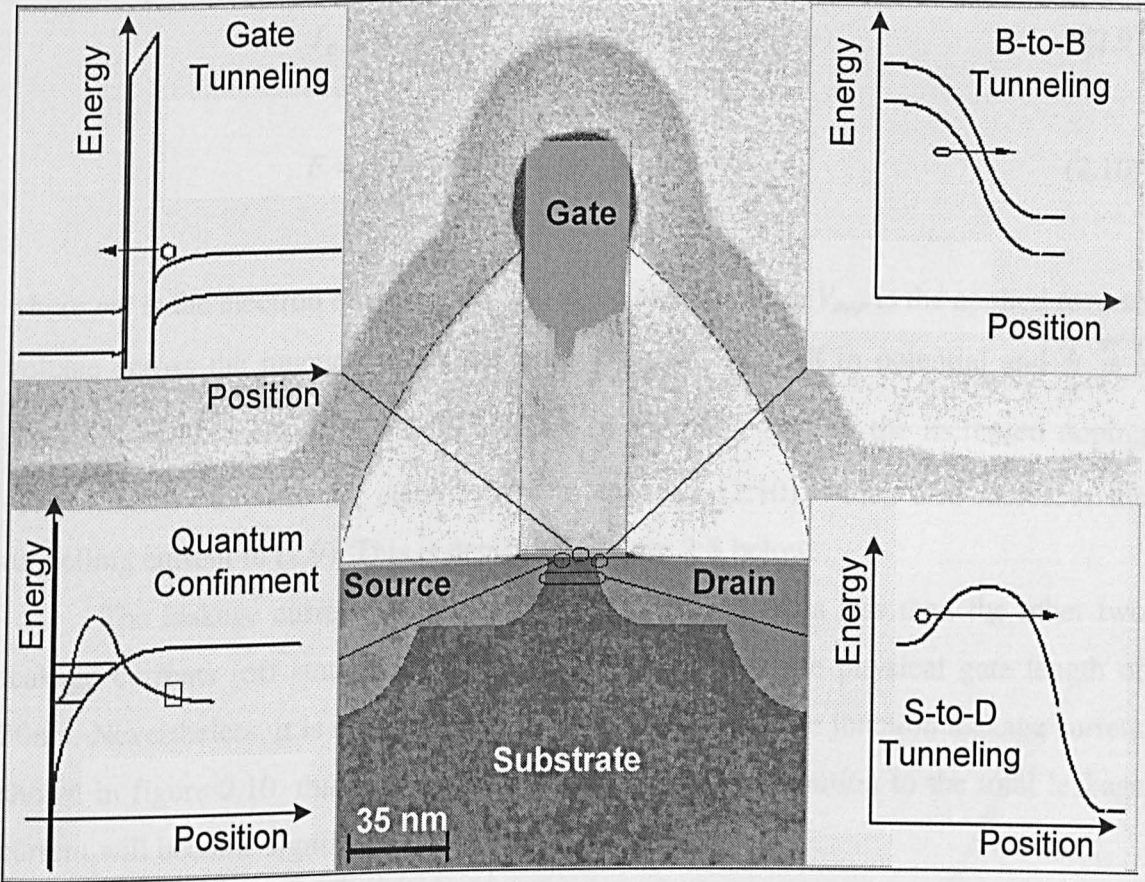


Figure 2:9 Visual illustrations of quantum effects near the Si/SiO2 interface: Reference [2.42]

2.3.1.1 Band to Band tunnelling

Band to band tunnelling (sometimes called Zener tunnelling) primarily occurs between the body and the drain of a MOSFET. The high channel doping that accompanies scaling results in a high electric field across the depletion layer at the reverse biased drain junction. The high electric field ($\sim 10^6$ V/cm) favours a parasitic leakage current associated mainly with the tunnelling of electrons from the valence band in the channel region to the conduction band in the drain [2.43] [2.8]. The tunnelling current density is approximated by:

$$J_{B-B} = \frac{\sqrt{2m^*} q^3 E V_{app}}{4\pi^3 \hbar E_g} \exp\left(-4 \frac{\sqrt{2m^*} E_g^{3/2}}{3qE\hbar}\right) \quad (2.9)$$

$$E = \sqrt{\frac{2qN_a(V_{app} + \psi_{bi})}{\epsilon_{si}}} \quad (2.10)$$

where m^* is the electron effective mass, E is the electric field, V_{app} is the applied reverse voltage across the junction, E_g is the energy gap, ψ_{bi} is built in potential and \hbar is a modified Planks constant ($\hbar = h/2\pi$). Due to transistor scaling the increased doping concentration increases the electric field in equation (2:10), which also increases the tunnelling current in (2.9). This is depicted in figure 2:8 below.

The leakage current due to band to band tunnelling is less than the other two leakage currents (off state and gate leakage) measured at the physical gate length of 30nm. Nevertheless, it is clear from the overall tendency of the junction leakage current shown in figure 2.10, that for smaller gate lengths its contribution to the total leakage current will become significant.

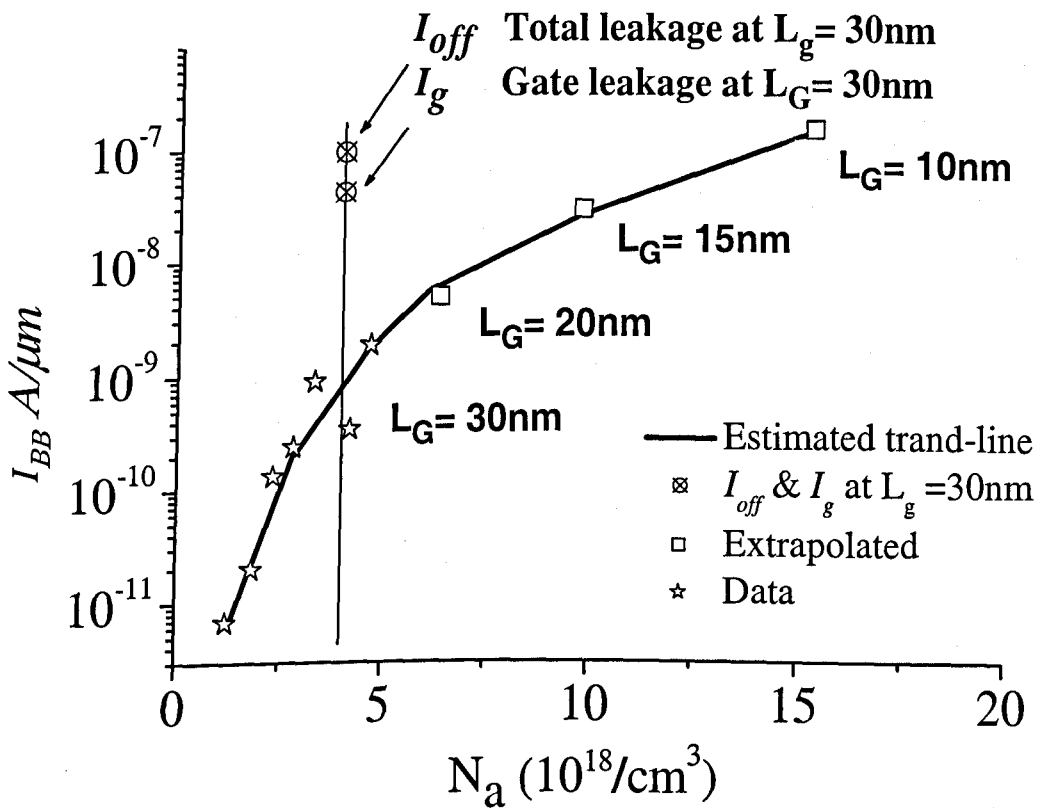


Figure 2.10 Junction leakage current due to band to band tunnelling (reference [2.44])

2.3.1.2 Direct gate oxide tunnelling

In order to achieve a desired current drive at a substantially low power supply voltages in sub 50 nm MOSFETs, aggressively scaled gate-dielectrics with equivalent oxide thickness (EOT) in the range of $t_{ox} \approx 1.5 - 0.5\text{nm}$ are required according to the latest edition of the ITRS. For such ultra-thin oxides the channel carriers can tunnel into the polysilicon gate through the gate-dielectric material. This process of electron or hole transmission through the dielectric barrier increases the gate leakage current exponentially with decreasing t_{ox} [2.45]. As a consequence of the increase in gate current, the overall off-state current is raised to an intolerable level for real circuit applications.

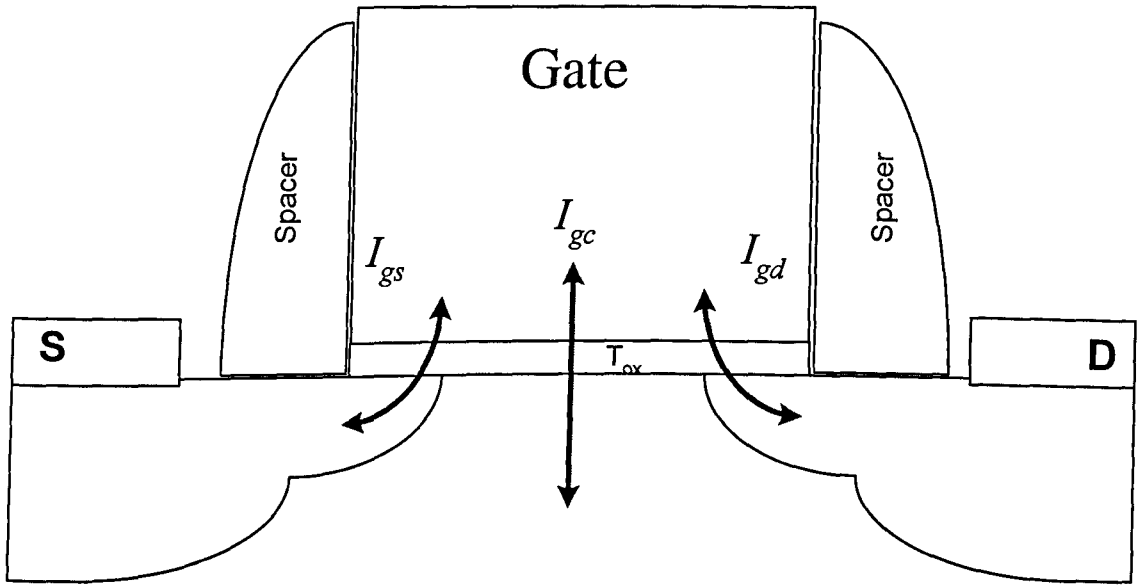


Figure 2:11 The overall gate tunnelling current is the sum all the components tunnelling current namely the source-to-gate, drain-to-gate and channel to gate currents (reference [2.46])

In the previous generations of MOSFET devices, the contribution of gate oxide leakage current to the overall leakage current has not been substantial. However, this is not the case for the present and the next generation of devices. As shown in figure 2.11, in addition to the tunnelling current from the channel I_{gc} , the fringing currents from the gate overlap with the source extension- I_{gs} and drain extension- I_{gd} also contribute to the total gate leakage current [2.46]. Bearing in mind that the main source of standby power drawn in CMOS circuits originates from the off-state current, substantial increase in off current due to direct tunnelling through the gate dielectric should be considered seriously during the design of devices or power optimization in circuits [2.47]. The application dependant power constraint as one of the main limitations of MOSFET scaling will be discussed in section 2.3.3

The cumulative gate tunnelling current density can be approximated by equation (2.4) [2.45]:

$$J_{DT} = \frac{4\pi q m_1 kT}{h^3} \int_0^{E_b} \tau_c(E) \ln \left[\frac{[\exp(E_{Fn1} - E_{C1} - E)/kT] + 1}{[\exp(E_{Fn2} - E_2 - E)/kT] + 1} \right] dE \quad (2.11)$$

where m_1 , and is the electron effective tunnelling mass E_{Fn1} the electron Fermi level, and E_{c1} is the bottom of conduction band in Si, and E_{Fn2} , and E_{c2} in the polysilicon region.

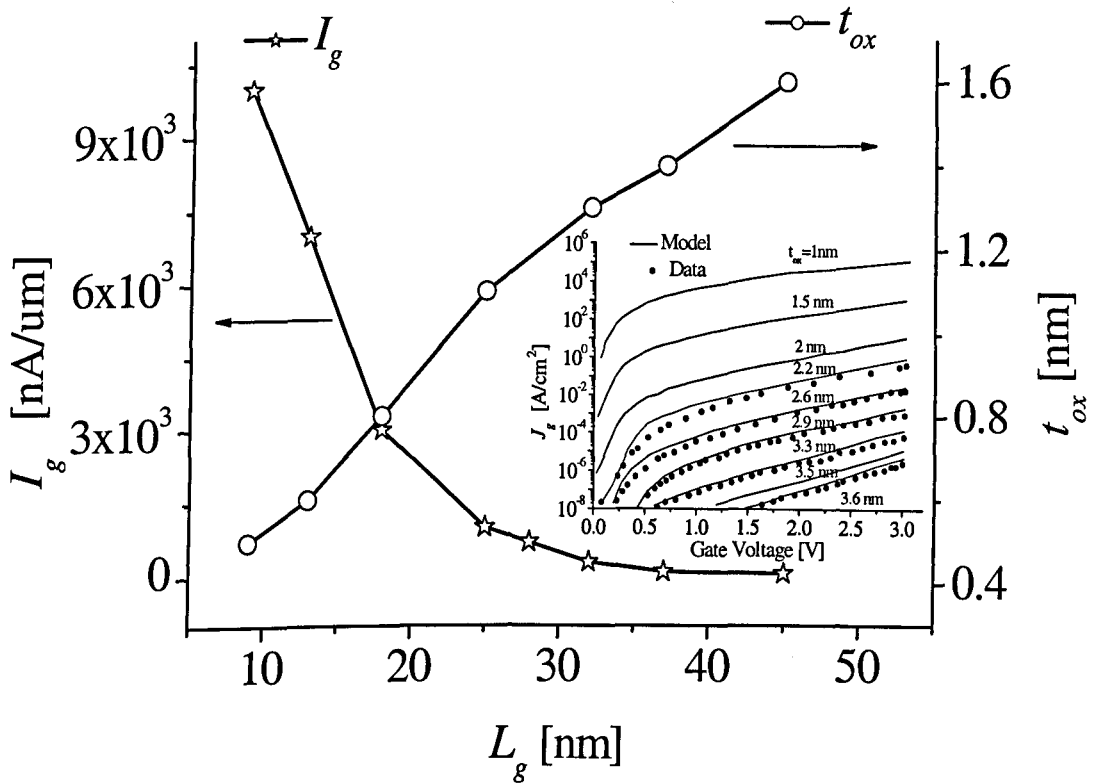


Figure 2:12 The relationship between the gate leakage current and the physical gate length. Inset: the gate current density for various oxide thicknesses as a function of gate voltage. Data from ITRS (2004 update) and the inset graph is adapted from reference. [2.48]

Figure 2:12 illustrates the relationship between gate length reduction and the gate oxide thickness which are required for the 90nm technology node and beyond. The figure also shows the increase of the gate leakage current density for various oxide thicknesses. The increase in leakage current is exponential [2.49]. For example, for gate lengths below 10 nm, and corresponding oxide thicknesses of 0.4 - 0.5, nm the leakage current reaches about 8-10 $\mu\text{A}/\mu\text{m}$ provided that SiO_2 is used as a gate dielectric material. This amount of leakage current can affect transistor integration in high performance digital systems where low power dissipation per area is a critical issue during the stand-by mode of operation.

The inset in figure 2:12 compares according to [2:48] the experimental data and the calculated gate tunnelling current density for a range of oxide thickness as a function

of the gate voltage. It illustrates well the exponential increase in the gate tunnelling with the reduction of the oxide thickness for the whole range of relevant gate voltages.

2.3.1.3 Source to drain tunnelling

The other possible cause of tunnelling current that can affect the operation of sub-10 nm MOSFETs is source-to-drain tunnelling. The proximity of the source and drain metallic-junctions may lead to quantum mechanical tunnelling that will increase the overall transistor leakage current. The effect of source to drain tunnelling current in an 8 nm MOSFET has been experimentally demonstrated by Kawaura *et al.* [2.50]. According to the technology roadmap such devices will be in production around year 2018.

The comparison of the subthreshold current and its temperature dependence made in [2.50] for two experimental devices with $L_g = 52$ nm and $L_g = 8$ nm shows that there is an increase in subthreshold leakage current and degradation of the subthreshold slope in an 8 nm transistor due to source to drain tunnelling. Although at room temperature the effect is negligible, the contribution of the source to drain leakage current at the lower temperatures increases. Due to the relatively low temperature sensitivity of the tunnelling current at low temperature when it dominates the subthreshold current, the subthreshold slope becomes temperature independent. It has also been shown in [2.50] that when source to drain tunnelling starts to play a big role the sub threshold slope becomes significantly degraded. This is also reported in [2.51] based on a density gradient device simulation approach.

2.3.2 Intrinsic parameter fluctuations

In addition to the investigation of the scaling properties of MOSFETs, the investigation of intrinsic parameter fluctuations in scaled MOSFETs is also considered as part of this project. Statistical simulation results of the effect of different sources of intrinsic parameter fluctuations, together with detailed discussions, are presented in chapter five. In this section, a background discussion on intrinsic parameter fluctuations in decananometre MOSFETs, and a brief review of some previous results are presented.

According to the latest update of the ITRS'04, transistors with gate length of 7nm are required for the 18 nm technology generation in the year 2018. The channel length of these devices is approximately 10-14 silicon atoms in span. The random number and position of dopants from device to device in such a small nominal volume of the crystalline lattice will introduce significant variation in key device parameters (such as threshold voltage and off-current) in any ensemble of devices. Gate line edge roughness (LER) and the oxide thickness variations are two other sources of intrinsic parameter fluctuations [2:52][2:53]. It has been shown, that for devices comparable to in size to the 35 nm MOSFET, interface roughness introduces significantly less fluctuations compared to the those from doping or LER, and will not be discussed in this work.

The intrinsic parameter fluctuations associated with random dopants in CMOS have been widely investigated over the last two decades. Initially the effect of random discrete dopants on V_T was studied by Hoeneisen and Mad in [2:54], which showed that a non uniform distribution of random dopants in the channel causes mismatch in the V_T of CMOS device. Hagiwara *et al.*, in the early 80's, investigated the effect of random dopants analytically and suggested a simple model to estimate V_T variation [2:55]. Furthermore, intrinsic parameter fluctuations in V_T , have been experimentally demonstrated and reported in [2:56] [2:57]. They have also been extensively studied theoretically based on statistical 3-D atomistic simulations by Asenov *et.al.*, [2:58 2:59], by Wang and Taur [2:60] and using a 3-D Monte Carlo approach by Frank *et al.*, [2:61].

Mizuno *et al.*, [2:57] have shown experimentally that the V_T fluctuations have a Gaussian distribution, which is a typical characteristic of random events. They also confirmed experimentally that the threshold voltage fluctuations increase with reduction of the channel length. The magnitude of the V_T fluctuations in their experimental device with a channel length of $L_{eff} = 500nm$ was about $\pm 3\sigma V_T = \pm 12mV$, which accounts for 60% of the overall fluctuations. Although this value is smaller than present fluctuations magnitude, its contribution to the overall fluctuations in devices is significantly high.

Extrapolation of these fluctuations, based on simulation studies, to decanano-meter device dimensions shows a very significant increase. For example, in the 35nm channel length MOSFET, the $\pm 3\sigma V_T$ increase up to $\pm 100mV$ [2:62] and are significantly larger for smaller gate length devices, as will be shown in chapter 5.

2.3.3 Power dissipation

Aside from the physical scaling discussed previously in this chapter, application dependant power dissipation becomes one of the major factors hampering the integration of the scaled devices and therefore limiting the usefulness of continued scaling [2.63].

From a VLSI circuits' application point of view, there are three power dissipation mechanisms, dynamic or switching power, short current generated power, and static power dissipations. These added together, give the total power dissipation, P_T , on integrated circuit:

$$P_T = \sum (P_D, P_{sh}, P_{st}) \quad (2.15)$$

Where $P_D = CV_{dd}^2 f$ is the dynamic (active) power dissipation due to the charging and discharging of the capacitive load on each one of the devices in the integrated circuit. The dynamic power dissipation takes place during transition from high to low logic level or vice versa. The second term in equation (2.15) $P_{sh} = V_{dd} I_{short}$ is the power dissipation due to occurrence of short circuit currents. This depends generally on the architecture of the circuit. Usually this is not the main component of power dissipation. The third component of the total power dissipation is the static power dissipation, $P_{st} = V_{dd} I_{leak}$, which occurs as a result of the cumulative leakage current contributed from all devices in the circuit.

By far the highest power consumption in present circuits comes from the dynamic activity of the devices in the circuit. The dynamic power dissipation is directly proportional to the square of the supply voltage. Although the power supply voltage is reducing as a result of device scaling, due to this quadratic relationship between P_D and V_{dd} , dynamic activity, still accounts for substantial power dissipation. However, the static power dissipation is gathering pace as the major power constraint in deeply scaled MOSFETs. This is illustrated in figure 2:13 which shows that the proportion of the static power dissipation in the 65 nm technology node circuits is increased compared to previous technology nodes.

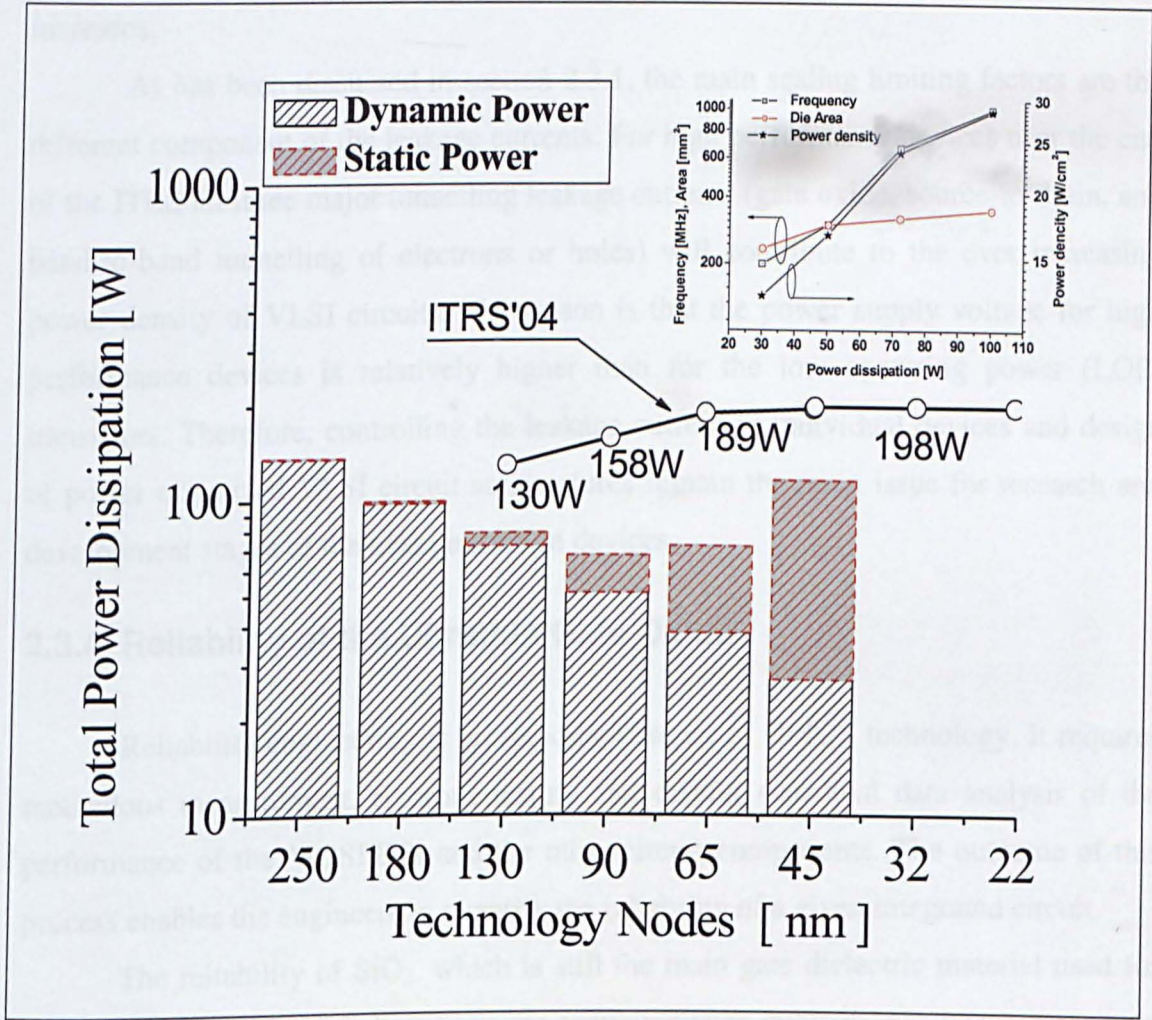


Figure 2:13 Total power dissipation as a function of technology node and an illustration of the increasing of the static power dissipation due to the significant increase in the off-state current as technology scales to 65nm and beyond. Data source: ITRS 2004 and references [2.64] [2.65]

For the 45 nm technology node the static power dissipation is expected to be more than 50% of the total power consumption (see figure 2:13). The ITRS requirement of maximum allowable power dissipation is also indicated in the graph and has been kept constant ($P_T=198W$) beyond the 65 nm technology node. The experimental data shown in the bar graph corresponds to 30 metres wide transistors for each technology generation [2.64].**

The inset in figure 2:13 depicts the data reported in [2.65] on the relationship between the power dissipation and the frequency. It also shows how the power density

** This is equivalent to 30 million transistors by considering the width of a single transistor is 1µm.

P_t/A is increasing significantly as the number of components integrated into a chip increases.

As has been discussed in section 2.3.1, the main scaling limiting factors are the different component of the leakage currents. For high performance devices near the end of the ITRS all three major tunnelling leakage currents (gate oxide, source-to-drain, and band-to-band tunnelling of electrons or holes) will contribute to the ever increasing power density of VLSI circuits. The reason is that the power supply voltage for high performance devices is relatively higher than for the low operating power (LOP) transistors. Therefore, controlling the leakage current in individual devices and design of power optimized VLSI circuit architectures remain the main issue for research and development stages of the next generation devices.

2.3.4 Reliability of the Ultrathin Gate Oxide

Reliability engineering is an important aspect of CMOS technology. It requires meticulous measurement, rigorous testing and careful statistical data analysis of the performance of the MOSFETs and the other circuit components. The outcome of this process enables the engineers to quantify the reliability of a given integrated circuit.

The reliability of SiO_2 , which is still the main gate dielectric material used for the fabrication of CMOS devices in the semiconductor industry, has been extensively studied over the past three decades. Most of the research has been concentrated mainly on trap creation (defect generation) mechanisms that degrade the gate oxide and eventually lead to its breakdown [2:66, 2:67, 2:68, 2:69]. For the sake of completeness, and to highlight in particular the reliability issues connected with possible breakdown of the aggressively scaled ultra-thin ($t_{ox} \leq 2nm$) gate dielectric materials, a brief discussion is presented in this sub-section.

Reliability in general is associated with the duration of time over which the device is operating at its full potential, or within some degree of tolerance. This period is called the "time-to-break down" t_{BD} which is commonly known as the product lifetime. With regard to the gate dielectric material reliability, it can also be associated with the charge build-up in time (related to the "charge-to-break down", $t_{BD} = Q_{BD}/J$, where J is gate leakage current) and Q_{BD} is given by equation (2.8) [2.70]

$$Q_{BD} = \frac{N_{BD}}{P_{gen}} \quad (2.14)$$

Where N_{BD} is the defect density which trigger breakdown and P_{gen} is the defect generation rate at which the defects are created and is given by (2.15)

$$P_{gen} = \left(\frac{\Delta J}{J} \right) \frac{1}{Q_{inj}} \quad (2.15)$$

Here $\Delta J/J_0$ is the relative change of the stress induced leakage current and Q_{inj} is the injected electron charge.

The main focus of this section is to relate the oxide thickness to reliability. However, equations 2.14 and 2.15 are not explicit function of the oxide thickness (t_{ox}). DiMaria *et al.*, [2.71] have found that the average density of states to breakdown (N_s^{BD}) decreases with decreasing t_{ox} . This means that fewer microscopic defects are enough to cause critical breakdown in very thin gate dielectric materials. Furthermore, Degraeve *et al.* [2.72], have investigated the dielectric material thickness dependence of the Weibull slope (β) defined below. They show that the Weibull slope becomes shallower with decreasing t_{ox} . The gate oxide thickness dependence of the Weibull slope is also reported in [2.73] using the cell-based statistical breakdown model. The Weibull function describes the cumulative probability of failure F , which is given by 2.16 [2.74]:

$$F(Q_{BD}) = 1 - \exp\left(-\left(\frac{Q_{BD}}{\alpha}\right)^\beta\right) \quad (2.16)$$

Equation (2.16) is often rewritten as in the form

$$\ln\left(\ln\left(1 - F(Q_{BD})\right)\right) = \beta \ln\left(\frac{Q_{BD}}{\alpha}\right) \quad (2.17)$$

where β is known as the shape factor of the distribution (often referred to as the Weibull slope) and α is the characteristic charge (or time) to breakdown at the reliability rate of 37% (or 63% failure percentage).

Figure 2:14 illustrates the relationships between the critical defect density (N_{BD}) and the oxide thickness. The critical defect density is strongly dependant on the oxide thickness and the amount of the charge to breakdown is highly dependant on the stress voltage.

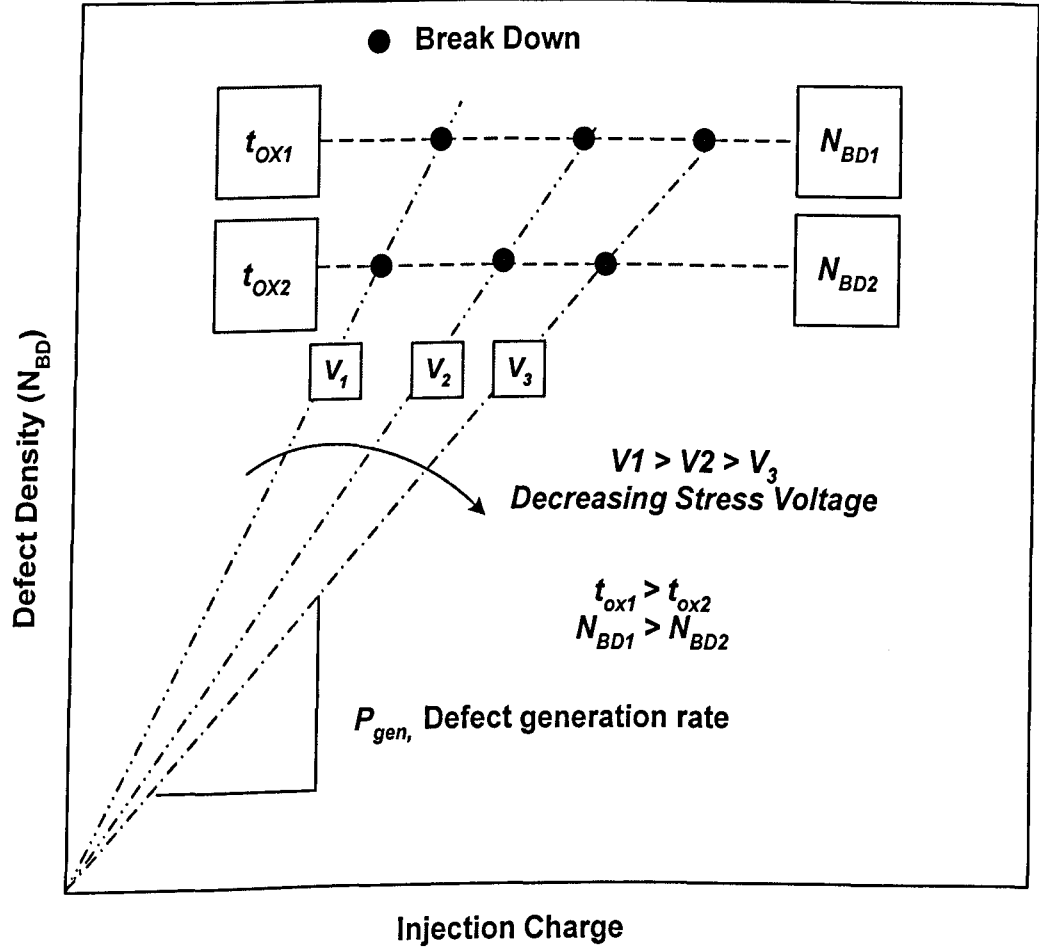


Figure 2:14 Schematic illustration of the relationship between the defect generation rate, defect density, oxide thickness and the stress voltage: Reference [2.70]

2.4 Chapter Summary

In this chapter we have reviewed Moore's law and the International Technology Roadmap for Semiconductor Technology. Both Moore's law and the ITRS have played a significant role in the advancement of the CMOS technology. In particular the road map has been an important guiding document that gives the direction for research and development for next generation devices. Apart from their role as a driving force behind the development of semiconductor technology as a whole, our emphasis was on their influence on MOSFETs scaling

The two most important scaling rules, 'constant field' and the 'generalised scaling,' have been discussed in detail. Based on the physical properties of short channel MOSFETs, the generalised scaling rule has been adopted for this work. It takes into account the 2D properties of the electric field and the retarded reduction of the supply voltage during the scaling process.

The limiting factors that hinder the progress of MOSFET scaling have also been discussed. Application dependant power constraints (including the leakage currents), the intrinsic parameter fluctuations, and the issue of the reliability of ultra-thin gate dielectric material are among the major scaling limiting factors that jeopardise Moore's law in the near future.

The main ultimate limitation to the scaling of transistors is the atomic structure of matter. A MOSFET cannot be smaller than an atom. The transistor post 2025, according to the IBM roadmap, will be reaching these limitations.

In the next chapter the simulation strategy and overall methodology used to investigate the scaling and intrinsic parameter fluctuation in nano CMOS devices are presented.

2.5 Chapter reference

- [2.1] G. E. Moore, "Cramming more components onto integrated circuits", *Electronics*, Vol. 38, number 8, April 19, 1965
- [2.2] J. Meindl, "Theoretical Practical and Analogical Limits in ULSI", *IEDM Tech. Digest*, IEEE press, pp. 8-13, 1983
- [2.3] G. M. Borsuk, and T. Coffey, "Moore's Law: A Department of Defence Perspective", *Defence Horizons*, Number 30, July 2003
- [2.4] International Technology Roadmap for Semiconductors, (SIA) <http://public.itrs.net>, SEMATECH, 2003 edition of ITRS
- [2.5] G. E. Moore, "Progress in Digital Integrated Electronics", *IEDM Tech. Digest*, IEEE press, pp. 11-13, 1975
- [2.6] R. Nair, "Effect of increasing chip density on the evolution of computer architecture", *IBM J. Res. & Dev.*, Vol. 46, No. 2/3, pp. 223-234, March/May 2002
- [2.7] P. P. Gelsinger, P. A. Gargini, G. H. Parker, A. Y. C. Yu, "Microprocessor circa 2000", *IEEE SPECTRUM*, pp. 43-47, October 1989
- [2.8] Y. Taur, and T. Ning, "Fundamentals of Modern VLSI Devices", Cambridge University Press, United Kingdom, 1998
- [2.9] E.J Novak, "Maintaining the benefits of CMOS scaling when scaling bogs down", *IBM J.Res & Dev.*, Vol. 46, NO. 2/3, 2002
- [2.10] G. Baccarani, M.R. Wordeman, and R.H. Dennard, "Generalised Scaling Theory and its Application to $\frac{1}{4}$ Micrometer MOSFET design," *IEEE Trans., Electron Devices*, Vol.31, pp. 452, 1984
- [2.11] J. R. Brews, *et. al.* "Generalised Guide for MOSFET Miniaturization, " *IEEE Electron Device letter*, Vol. EDL-1, pp. 2-4, 1980
- [2.12] S. Inab, K. Okano, S. Matsuda, M. Fujiwara, A. Hokazono, K. Adachi, K. Ohuchi, H. Suto, H. Fukui, T. Shimizu, S. Mori, H. Oguma, A. Murakoshi, T. Itani, T. Iinuma, T. Kudo, H. Shibata, S. Taniguchi, M. Takayanagi, A. H. Oyamatsu, K. Suguro, .Y. Katsumata Y. Toyoshima, and H. Ishiuchi, "High Performance 35 nm Gate Length CMOS With NO Oxynitride Gate Dielectric

-
- and Ni Salicide", *IEEE Trans. Electron Devices*, Vol. 49(12), pp. 2263 – 2270, 2002
- [2.13] D. J. Frank, R. H. Dennard, E. Nowak, P. M. Solomon, Y. Taur, and H-S. P. Wong, "Device Scaling Limits of Si MOSFETs and Their Application Dependencies", *Proceedings of The IEEE*, Vol. 89, No. 3, pp. 259-288, March 2001
- [2.14] R.H. Dennard, F. H. Gaensslen, L. Khun and, H. N. Yu, "Design of micron switching devices", *Presented at IEDM*, Washington D.C., December 1972
- [2.15] R.H. Dennard, F. H. Gaensslen, Hw A-Nien Yu, V. L. Rideout, E. Bassous, and A. R. LeBlanc, "Design of Ion-Implanted MOSFET's with Very Small Physical Dimensions", *IEEE Journal of Solid-State Circuits*, Vol. SC-9, No. 5, pp. 256-268, 1974
- [2.16] N. D. Arora, J. R. Hauser, and D. J. Roulston, "Electron and Hole Mobilities in Silicon as a Function of Concentration and Temperature," *IEEE Trans Electron Devices*, Vol. Vol. 29, pp. 292-295, Feb. 1982.
- [2.17] K. K. Ng, S. A. Eshraghi, and T. D. Stanik, "An Improved Generalised Guide for MOSFET Scaling", *IEEE Trans Electron Devices*, Vol. 40(10), pp. 1895-1897, 1993
- [2.18] T. Skotnicki, "Heading for decanano CMOS- Is navigation among the iceberg still a viable strategy?", *Proceedings of ESSDERC2000*, pp. 19-33, Sep. 2000
- [2.19] T.Skotnicki, C. Denat, P. Senn, G. Merckel, and B. Hennion, "A new analog/digital CAD model for sub-halfmicron MOSFETs", *IEDM Tech. Dig.*, pp. 165 - 168, December 1994
- [2.20] D. J. Frank, "Power-constrained CMOS scaling limits", *IBM J. Res. & Dev.*, Vol. 46, No. 2/3, pp. 235-244, March/May 2002,
- [2.21] C-T. Sah, "Evolution of the MOS Transistor –From Conception to VLSI", *Proceedings of the IEEE*, Vol. 76 (10), pp. 1280-1326, 1988
- [2.22] J. R. Burns, "Switching response of complimentary MOS transistor logic circuit", *RCA Review*, Vol. 25, pp. 627-661, December 1964
- [2.23] A.K. Rapp, L. P. Wennick, H. Borkan and K. R. Keller, "Complementary-MOS Integrated binary Counter", *1967 International solid-state circuit conference, Digest of technical papers*, pp.52-53, 1967

-
- [2.24] T. D. Kamins, "Preparation and Polycrystalline-Silicon Films ", in Handbook of Semiconductor Silicon Technology, Edited by W. C. O'Mara, Robert B. Herring, and L. P. Hunt, Noyes Publications, pp. 640-730, 1990
 - [2.25] K. Saito, T. Morose, S. Sato, and U. Harada, "A new short channel MOSFET with lightly doped drain", *Danshi Tsushin Rengo Taikai*, pp. 220, April 1979
 - [2.26] S. O., Paul, J. Tsang, W. W. Walker, L. D. Critchlow, and J. F. Shepard, "Design and Characteristics of the Lightly Doped Drain-Source (LDD) Insulated Gate Field-Effect Transistor", *IEEE J. Solid-State Circuit*, Vol.SC-15(4), pp. 424-432, 1980
 - [2.27] P. J. Tsang, S. Ogura, W. W. Walker, J. F. Shepard and L. D. Critchlow, "Fabrication of High-Performance LDDFET's with Oxide Sidewall Spacer Technology ", *IEEE J. Solid-State Circuit*, Vol.SC-17(2), pp. 220-432, 1982
 - [2.28] K.-H. Oh, Z. Yu, and R. W. Dutton, "A Bias Dependant Source/Drain Resistance Model in LDD MOSFET Devices for Distortion Analysis", *6th International Conference on VLSI and CAD: ICVC'99*, pp. 190-193, 1999
 - [2.29] B. L. Crowder and S. Zirinsky, "1-micron MOSFET VLSI technology. VII - Metal silicide interconnection technology: A future perspective", *IEEE J. Solid-state Circuits*, SC-14, pp. 291-293, 1979
 - [2.30] G. G. Shahidi, D. A. Antoniadis, and H. I. Smith, "Indium channel implant for improved short-channel behaviour of sub micrometer NMOSFET's," *IEEE Trans. Electron Devices*, vol. 14, p. 409, Aug. 1993.
 - [2.31] I. De and C. M. Osburn, "Impact of Super-Steep-Retrograde Channel Doping Profiles on the Performance of Scaled Devices", *IEEE Trans. Electron Devices*, vol. 46, p. 1711-1717, 1999
 - [2.32] Y. Taur, C. H. Wann, and D. J. Frank, "25nm CMOS Design Consideration", *IEDM'98 Technical Digest*, pp. 789-792, 1998
 - [5:33] S. Thompson, *et al.*, "A 90nm technology featuring 50nm strained silicon channel transistors, 7 layers of Cu interconnects, low k ILD and 1um 2SRAM cell ," 2002 IEDM Digest, pp. 21-64, Dec. 2002
 - [2:34] W. M. Siu, "Forward", *Intel Tech. J.* Vol 8 (1), 2004
 - [2:35] K. Goto, *et al.*, *IEDM Tech Dig*, IEDM'03, pp. 623, 2003

-
- [2.36] B. Yu, H. Wang, A. Joshi, Q. Xiang, E. Ibok, M.-R. Lin, "15nm Gate length Planer CMOS Transistor", *IEDM Tech Dig*, IEDM'01, pp. 937 - 939, 2001
 - [2.37] F. Boeuf, T. Skotnicki, S. Monfray, C. Julien, D. Dutartre, J. Martins, P. Mazoyer, R. Palla, B. Tavel, P. Ribot, E. Sondergard, and M. Sanquer, "16nm planar NMOSFET manufacturable within state-of-the-art CMOS process thanks to specific design and optimisation", *IEDM Tech Dig*, IEDM'01, pp. 637 - 640, 2001
 - [2.38] A. Hokazono *et al.*, "14 nm Gate Length CMOSFETs Utilizing Low Thermal Budget Process with Poly-SiGe and Ni Salicide", *IEDM Tech Dig*, IEDM'02, pp. 639 - 642, 2002
 - [2.39] N. Yasutake, *et al.*, "A hp22nm Node Low Operating Power (LOP) Technology with Sub-10nm Gate Length Planar Bulk CMOS Devices", Digest of Technical Papers, 2004 Technology Symposium, pp. 84, 2004
 - [2.40] B. Doris, *et al.*, "Device Design Considerations for Ultra-Thin SOI MOSFETs", *IEDM Tech Dig*, IEDM'03, pp. 631 - 634, 2003
 - [2.41] H. Wakabayashi, "Transport Properties of Sub-10-nm Planar-Bulk-CMOS Devices", *IEDM Tech Dig*, IEDM'04, pp. 429 - 432, 2004
 - [2.42] A. Asenov, J. R. Watling, A. R. Brown, D. K. Ferry, "The use of Quantum Potential for Confinement and Tunnelling in Semiconductor devices", *Journal of Computational Electronics*, 1:503-513 2002
 - [2.43] S. M. Sze, Semiconductor devices : Physics and technology, John Wiley & Sons, INC, 2nd Edition, 2001
 - [2.44] B. Doyle *et al.*, "Transistor Elements for 30nm Physical Gate Length and beyond," *Intel Technical Journal*, Vol. 06, Issue 02, pp. 42, 2002
 - [2.45] S. -Lo, D. Buchannen, Y. Taur, and W. Wang, "Quantum-Mechanical modelling of Electron Tunnelling Current from Inversion Layer of Ultra-Thin-Oxide nMOSFETs", *IEEE Electron Device Lett.*, Vol. 18, pp. 209-211, 1997
 - [2.46] C.-H. Choi, K.-Y. Nam, Z. Yu, and R. W. Dutton, "Impact of Gate Direct Tunneling Current on Circuit Performance: A Simulation Study", *IEEE Trans, Electron Devices*, Vol. 48(12), pp. 2823-2829, 2001
 - [2.47] D. Lee, D. Blaauw, and D. Sylvester, "Gate Oxide Leakage Current Analysis and Reduction for VLSI Circuits" *IEEE Trans. Very Large Scale Integration(VLSI) Systems*, Vol. 12(2), pp. 155-166, 2004

-
- [2.48] Y. Taur, "CMOS design near the scaling limit" *IBM J. Res. Develop*, Vol. 46(2/3), pp. 213-222, 2002
 - [2.49] S. -H. Lo, D. A. Buchanan, and Y. Taur, "Modeling and characterization of quantization, polysilicon depletion, and direct tunnelling effects in MOSFETs with ultra thin oxides", *IBM J. Res. Develop*, Vol. 43(3), pp. 327-337, 1999
 - [2.50] H. Kawaura, T. Sakamoto, and T. Baba, "Observation of source-to-drain direct tunneling current in 8 nm gate electrically variable shallow junction metal-oxide-semiconductor field-effect transistors", *Applied Physics Letter*, Vol. 76, NO. 25, pp. 3810 - 3812, 2000
 - [2.51] J. R. Watling, A. R. Brown, and A. Asenov, "Can the Density Gradient Approach Describe the Source-Drain Tunneling in Decanano Double-Gate MOSFETs", *Journal of Computational Electronics*, 1:289-293 2002
 - [2.52] A. Asenov, S. Kaya, and J. H. Davies, "Intrinsic Threshold Voltage fluctuations in Decanano MOSFETs Due to Local Oxide Thickness Variations", *IEEE Tran. Electron Devices*, Vol. 49 (1), pp. 112-119, 2002
 - [2.53] A. Asenov, S. Kaya, and A. R. Brown, "Intrinsic Parameter Fluctuations in Decananometer MOSFETs Introduced by Gate Line Edge Roughness", *IEEE Tran. Electron Devices*, Vol. 50(5), pp. 1254-1260, 2003
 - [2.54] B. Hoeneisen and C. A. Mad, "Fundamental Limitation in Microelectronics –I. MOS Technology", *Solid-State Electronics*, Vol. 15, pp. 819-829, 1972
 - [2.55] T. Hagiwara, K. Yamaguchi, and S. Asai, "Threshold voltage variation in very small MOS transistors due to local dopant fluctuations," *Proc. Symp. VLSI Technol., Dig. Tech. Papers*, 1982, pp. 46–47.
 - [2.56] R. Kadaba, R. Lakshmikumar, A. Hadaway, and M. A. Copeland, "Characterization and Modeling of Mismatch in MOS Transistors for Precision Analog Design", *IEEE journal Of Solid-State circuits*, Vol. Sc-21, No. 6, pp. 1057, 1986
 - [2.57] T. Mizuno, J -I. Okamura, and A. Toriumi, "Experimental Study of Threshold Voltage Fluctuation Due to Statistical Variation of Channel Dopant Dopant Number in MOSFET's", *IEEE Trans, Electron Devices*, Vol. 41(11), pp. 2216-2221, 1994

-
- [2.58] A. Asenov, "Statistically reliable 'Atomistic' simulation of Sub 100nm MOSFETs", in *Simulation of Semiconductor Process and Devices*, edited by K. De Meyer and S. Biesemans, Springer, Wien New York, pp. 223-226, 1998
 - [2.59] Asen Asenov, Andrew R. Brown, John H. Davies, Savas Kaya, and Gabriela Slavcheva, "Simulation of Intrinsic Parameter Fluctuations in Decananometer and Nanometer-Scale MOSFETs", *IEEE Trans, Electron Devices*, Vol. 50(9), pp. 1837-1852, 2003
 - [2.60] H. S. Wong, and Y. Taur, "Three dimensional 'atomistic' simulation of discrete random dopant effects in sub-0.1 μm MOSFETs", *Proc. IEDM'93 dig. Tech. Papers*, pp. 705, 1993
 - [2.61] D. J. Frank, Y. Taur, M. Jeong, and H-S P. Wang, "Monte Carlo Modeling of Threshold Variation due to Dopant Fluctuation", *Symposium on VLSI Circuits Digest of Technical Papers*, pp. 171-172, 1999
 - [2.62] F. adamu-Lema , G. Roy, A. R. Brown, A. Asenov, and S. Roy, "Intrinsic Parameter Fluctuations in Conventional MOSFET's at the Scaling Limit: A Statistical Study", *Abstracts to IWCE10*, pp.44-45, 2004
 - [2.63] D. J. Frank, "Power constraint CMOS scaling limits", *IBM J. RES. DEV.* Vol. 46 (2/3), pp. 235-244, 2002
 - [2.64] B. Doyle *et al*, "Transistor Elements for 30nm Physical Gate Lengths and Beyond", *Intel Technology Journal*, Vol. 6(2), pp. 42-54, 2002
 - [2.65] T. Mudge, "Power: A first-Class Architect Design Constraint ", *IEEE Computer*, Vol. 34, pp 52-58, April 2001
 - [2.66] A. Berman, "Time Zero Dielectric reliability Test by a Ramp Method"
 - [2.67] J. M. Aitken, and D. R. Young, "Electron trapping by radiation-induced charges in MOS devices", *J. Appl. Physics*, Vol. 47, pp. 1196-1198, 1976
 - [2.68] D. J. DiMaria and J. W. Stasiak, "Trap creation in silicon dioxide produced by hot electrons", *J. Appl. Physics*, Vol 55(6), pp. 2342-2356, March 1989
 - [2.69] R. Degraeve, "Oxide Reliability Issues", in *High Dielectric Constant Materials, VLSI MOSFET Application*, Springer, pp. 91-120, 2005
 - [2.70] D. A. Buchnaan, "Scaling gate dielectric: Materials, integration, and reliability", *IBM J. Res. Develop*, Vol. 43(3), pp. 245-264, 1999

-
- [2:71] D. J. DiMaria and J. H. Stathis, "Explanation for the oxide thickness dependence of breakdown characteristics metal-oxide- semiconductor structures", *Appl. Physics Lett.*, Vol. 70(20), pp. 2708-2710, May 1997
- [2:72] R. Degraeve, G. Groeseneken, R. Bellens, M. Depas, and H. E. Maes, "A consistent model for the thickness dependence of intrinsic breakdown in ultra-thin oxides", *IEDM Tech. Dig.*, pp. 863 - 866, 1995
- [2:73] D. J. DiMaria, and J. H. Stathis, "Explanation for the oxide thickness dependence of breakdown characteristics of metal-oxide-semiconductor structures", *Appl. Phys. Lett.*, Vol. 70, pp. 2708-10, 1997
- [2:74] E. Y. Wu. And R. -P. Vollertsen, "On the Weibull shape Factor of Intrinsic Breakdown of Dielectric Films and Its Accurate Experimental Determination- Part I: Theory, Methodology, Experimental Techniques", *IEEE Trans. Electron Devices*, Vol. 49(12), 2002.

3

Simulation Methodology

This chapter discusses the simulation methodologies which have been adopted during the course of this research programme. The first section discusses the principal concepts and the overall structure of the integrated atomistic process and device simulation strategy implemented with Technology Computer Aided Design (TCAD) device modelling tools.

Detailed descriptions of the models involved in the process simulation of device fabrication sequences; including ion implantation, deposition, etching, diffusion and oxidation is the subject of the second section. The last section in this chapter provides an overview of device simulation methodology, by reviewing some of the physical models and simulation approaches used in this work.

All the simulation techniques discussed in the three sections of this chapter have been calibrated and validated against real 35nm physical gate length MOSFETs fabricated and reported by Toshiba [3.1]. They have been used subsequently to scale the transistors according to the requirements of the ITRS (2003 edition) and to study the impact of intrinsic parameter fluctuations on such scaled devices.

3.1 TCAD and Integrated process & device simulation

The state of the art in modelling and simulation covers a wide range of specialist areas related to the entire process of integrated semiconductor device manufacturing. The following are some of the main areas of modelling and simulation [3.2] typically covered by state of the art TCAD simulation tools:

- Front-end process modelling that characterizes the essential steps for device manufacturing,
- Device modelling and simulation in order to understand how the devices operate by studying their electrical characteristics and intrinsic physical properties,
- Compact modelling to perform circuit level device simulation
- Circuit modelling which simulates the behaviour of the integrated circuit as a module,
- Modelling of equipment to study the general features of the deposition, etching and other manufacturing processes.

However, a detailed discussion of all these physical models and simulation methodologies is not the aim of this section. Therefore, the emphasis has been limited to cover only the front end process and device modelling related to the aim and objectives of this thesis.

3.1.1 The role of TCAD in advanced device design and characterisation

As the processing power of computer technology doubles roughly every two years, according to Moore's law [3.3], the requirements of TCAD for advanced device modelling, simulation and design expands considerably. TCAD is the synthesis of process, device and circuit simulations using state of the art computer technology. One of the most important advantages of using TCAD simulation and modelling is to produce deep insight and understanding of the physical process involved in the operation of the modern semiconductor devices.

Another dimension of the importance of TCAD in industry and research environments is its role in reducing the time required for the development of new generations of devices. This has an enormous implication on the over-all cost associated with scaling and new product development. With the present day's availability of supercomputer technology it is possible to have a virtual simulation laboratory that is capable of predicting the behaviour of new devices, and their electrical parameters, resulting from a particular fabrication sequence with a reasonable degree of accuracy [3.4]. This capability permits the design of an optimized prototype device and technology in a relatively short period of time and with reduced development cost.

In addition, TCAD modelling and simulation is also a vital tool in failure analysis, trouble-shooting and inverse modelling. By incorporating statistical functions into the TCAD package, it is possible to perform failure analysis and reliability tests. In the inverse modelling process, the initial point is a device with a poorly-defined structure, but known electrical parameters.

Using device simulation and calibration, one can deduce the realistic device structure so that it can be used for further development purposes or scaling studies. This will also be an important tool in the investigation of new materials and their impact on device performance. Using simulation tools, one can perform not only a preliminary analysis on the new material's (for example high- κ dielectrics), reproducibility, reliability and interface properties, but could also study their impact on the behaviour of the new device.

The ever decreasing dimensions of CMOS devices, (now down to the sub 100nm technology nodes) together with the ever increasing complexity of manufacturing technology [3.5] highlight the necessity of process and device simulation more than ever. Technologically demanding and financially costly experimental procedures can now be modelled and thoroughly analysed in a relatively short time period.

All the benefits mentioned above emphasise the importance of TCAD simulation as a tool in the field of device research and development. Despite all the advantages in the development of new generation devices, TCAD has its own drawbacks. There is always some degree of a mismatch between the measured data and the simulated values. It is also possible to obtain completely misleading results from device or process simulations.

The reasons could be the use of unphysical models, inappropriate simulation techniques, and inaccurate default material and transport parameters. Moreover, most simulation models are to some degree empirical, and require numerical calibration to experimental data at each technology node. This means that simulation accuracy cannot automatically be extended to the next generation of devices and technology nodes. This is particularly true for the present generation of sub-100 nm devices where the short channel effect, non-equilibrium transport, and quantum mechanics govern their operation.

It is also very important to underline that computer simulation is not the final word in overcoming the challenges in developing present and future technology

generations. It can only be used as a probing instrument and predictive tool that facilitates the environment for advanced MOSFET design and fabrication. Definite analysis and conclusions still require confirmation, obtained from experimental results.

3.1.2 Integrated process and device simulation in 3D

The underlying philosophy behind the integrated process and device simulation of decanano-meter MOSFETs [3.6] is the systematic and methodological selection and linking of various types of semiconductor TCAD tools efficiently. It is a “divide and conquer” strategic approach which applies the best available models and tools for designing and understanding scaled devices of succeeding generations of semiconductor technology.

One of the main aims of this PhD work is to study the impact of different sources of intrinsic parameter fluctuations on “real devices” created by the simulation of a realistic process sequence. The nature of the fabrications requires a complete 3D process and device simulation and is an example of the most demanding application of TCAD tools.

In order to make 3D process model as realistic as possible, a well behaved 35 nm physical gate length MOSFET reported by Toshiba [3.1] has been adopted as a bench mark device for this work. Here we will use the simulation of this device as an example of the 3D integrated device simulation approach. Initially, comprehensive 3D process simulation of the 35 nm device based on the real device fabrication sequence was carried out with the 3D process simulator, Taurus, in order to deduce the device geometry and doping profile. This information was subsequently exported to the device part of the simulator, which gives electrical characteristics. These were then compared with the experimental data in order to adjust the transport parameters and validate the simulators.

The overall methodology of integrated process and device simulation used in this work is illustrated on figure 3:1. The scaling of the 35 nm MOSFETs to the smaller dimensions which correspond to the future technology nodes was then performed using a continuous doping Taurus process and device simulator. The impact of Line Edge Roughness (LER) was also studied using continuous doping effects.

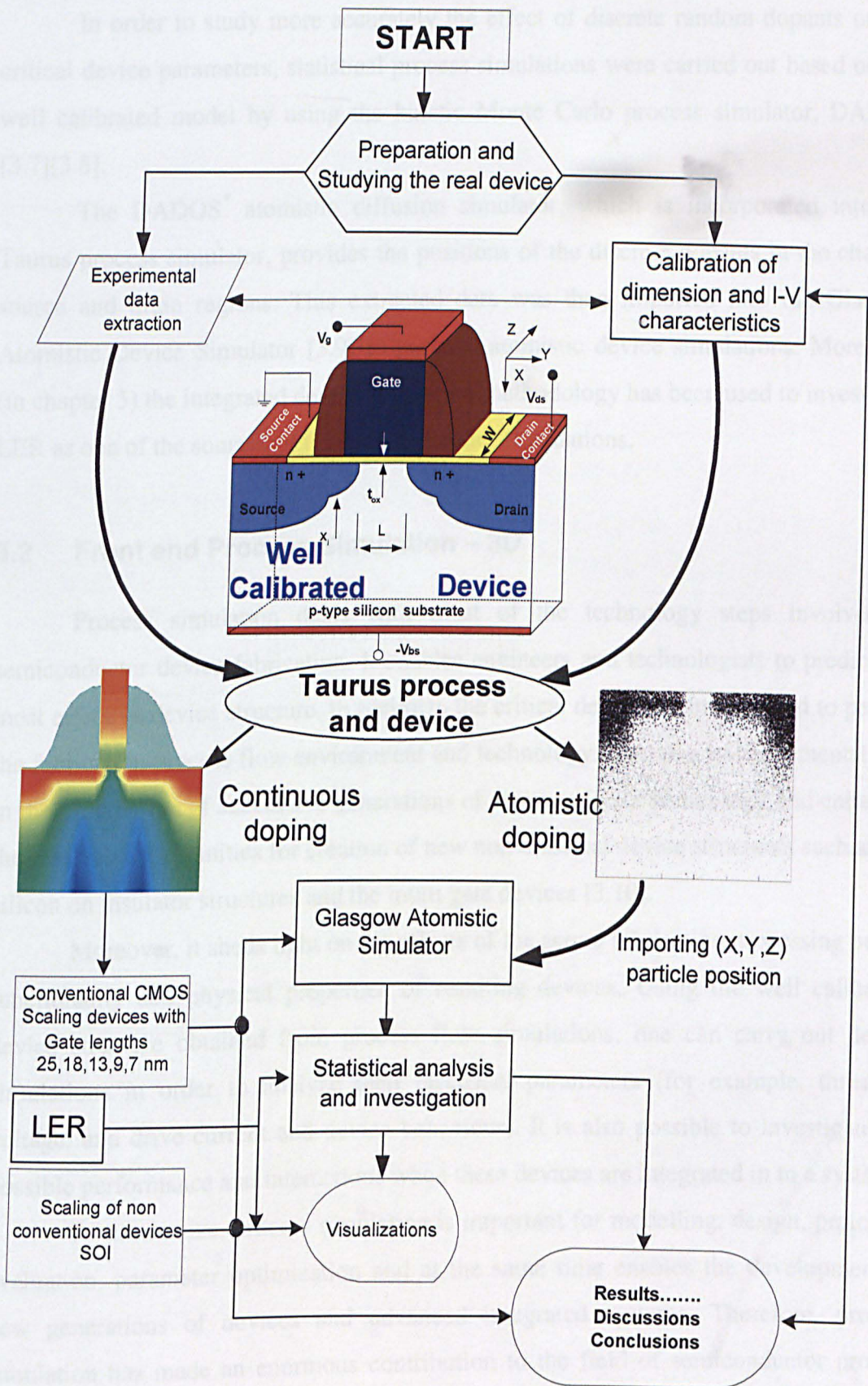


Figure 3:1 Flow chart that illustrate the integrated process and device simulation methodology.

In order to study more accurately the effect of discrete random dopants on the critical device parameters, statistical process simulations were carried out based on the well calibrated model by using the kinetic Monte Carlo process simulator, DADOS [3.7][3.8].

The DADOS* atomistic diffusion simulator, which is incorporated into the Taurus process simulator, provides the positions of the discrete dopants in the channel source and drain regions. This extracted data was then imported into the Glasgow Atomistic Device Simulator [3.9] to perform atomistic device simulations. Moreover, (in chapter 5) the integrated device simulation methodology has been used to investigate LER as one of the sources of intrinsic parameter fluctuations.

3.2 Front end Process Simulation – 3D

Process simulation deals with most of the technology steps involved in semiconductor device fabrication. It enables engineers and technologists to predict the most effective device structure, to optimize the critical device parameters and to perfect the fabrication process flow environment and technology. It is also an instrumental tool in the progression of successive generations of semiconductor technology and enhances the research opportunities for creation of new non-classical device structures such as the silicon on insulator structures and the multi gate devices [3.10].

Moreover, it sheds light on the effects of the actual fabrication processing on the functionality and physical properties of resulting devices. Using the well calibrated device structure obtained from process flow simulations, one can carry out device simulations in order to analyze their electrical parameters (for example, threshold voltage, and drive current and device behaviour). It is also possible to investigate the possible performance and interactions when these devices are integrated in to a system.

To summarize, process simulation is important for modelling, design, prototype evaluation, parameter optimization and at the same time enables the development of new generations of devices and advanced integrated systems. Therefore, process simulation has made an enormous contribution to the field of semiconductor process integration.

* More discussions on DADOS is given in section 3.2.5

As mentioned in the previous sections, the research presented in the thesis is underpinned by data extracted from realistic MOSFETs. The calibration of the simulation methodology in respect of these devices, which involves a painstaking process simulation, is a vital ingredient of the overall simulation methodology. Taking this into consideration, a step-by-step calibration strategy (described in detail in chapter 4) was employed to perform process and device simulations.

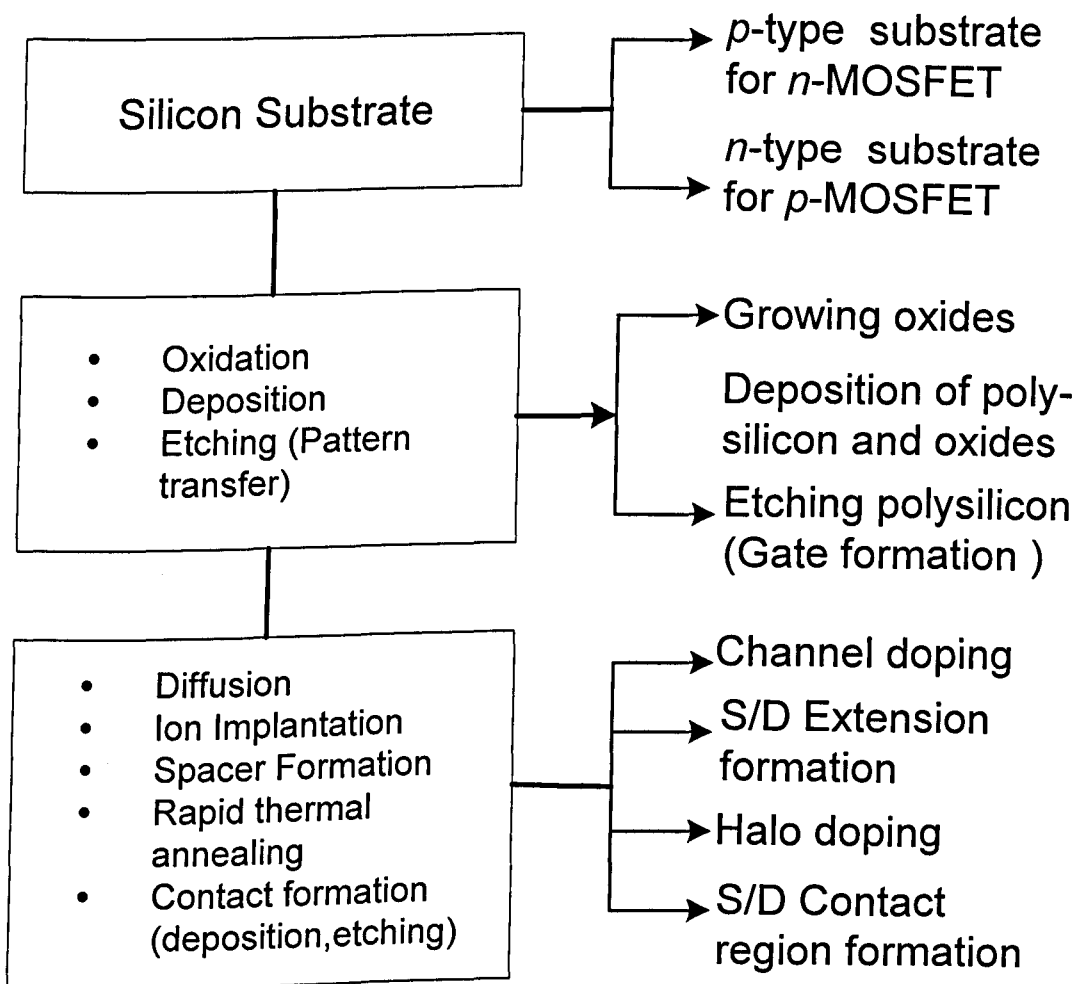


Figure 3:2 Summaries of CMOS device fabrication steps

The starting point of this process was the identification of the process steps on corresponding parameters for Toshiba's 35 nm MOSFETs. This includes the total ion dose of the well and channel implantation; the doping concentrations in the source and drain regions as well as the halo doping conditions; the annealing temperatures and the time and implantation parameters for all dopants. In addition to these preliminary input

data required to run process simulations, it is also equally important to extract the device dimensions such as the gate length, oxide thickness, the source/drain shallow junction depth and to define a computationally reasonable size of simulation domain. Such extraction is mainly based on high resolution TEM images. Once the above information has been obtained, 'Taurus Process and Device' simulation software [3.11] was employed to perform a realistic device process flow simulation. Taurus Process and Device simulation is a commercial TCAD tool enhanced by a 3D simulation capability with a wide range of physical simulation model options.

In the course of this research project, the main fabrication processes steps involved in the device process simulation sequences are: ion implantation, deposition, etching, and diffusion as a result of rapid thermal annealing (RTA). The steps are summarised in figure 3.2. Although this section does not attempt to give a comprehensive review of these device fabrication steps, a general discussion of the processes involved with the process flow steps included in the integrated device simulation methodology will be presented in the following sections.

3.2.1 Implantation – The Introduction of impurities in to Si crystal

A pure, intrinsic silicon crystal has a poor electrical conduction due to a low intrinsic carrier concentration ($1.4 \times 10^{10} \text{ cm}^{-3}$) at 300K [3.12]. However, it is possible to significantly change the local electrical conductivity or the conductivity type of the silicon crystal by introducing a controlled amount of impurity atoms (dopants) [3.13]. Currently, there are two ways to achieve this: the first method is the diffusion of dopant atoms from gaseous, liquid, or solid sources and the second is via ion implantation into the host semiconductor. In order to be consistent with the Toshiba MOSFETs fabrication sequence, this discussion concentrates on ion implantation.

The process of selectively incorporating extrinsic ions into semiconductor materials through surface bombardment of high velocity (and energy) is known as ion implantation [3.14]. It has been used for over two decades in silicon integrated process technology.

Ion implantation has several important advantages over the process of diffusion of dopant atoms into semiconductors. One of the main advantages is that there is a great deal of control over the amount of introduced impurities and over the impurity profile, *i.e.* the ability to locate the impurities at the desired depth of the host semiconductor [3.15] [3.16]. Another advantage is the possibility of using various types of ions which are otherwise difficult to introduce by diffusion [3.17]. The capability of performing ion implantation within a room temperature environment (reducing the effects of out-diffusion and the undesirable spread of the intended profile) and in a relatively short period of time is also another advantage.

Generally, elements in group **III** (old group nomenclature) (trivalent impurities) and group **V** (pentavalent impurities) of the periodic table have been used as dopants in silicon crystals by semiconductor manufacturers for decades. The chemical properties of group-**III** elements initiate a *p*-type conductivity (generally known as acceptors) by creating hole or electron deficiency. On the other hand group-**V** elements initiate *n*-type conductivity (donors), by providing an extra electron. Accordingly, indium for the *n*-MOSFET and arsenic for the *p*-MOSFET channel doping have been used to manufacture the 35nm Toshiba MOSFETs, which are at the focus of this research.

Boron and $(\text{BF}_2)^+$ are the most commonly used impurities in semiconductors in the past as well as at the present time. Recently, however, indium is beginning to replace these materials as recent research has shown that indium, due to its low diffusion coefficient, is a good choice to form the super steep retrograde channel (SSRC) doping profiles required beyond the $0.18\mu\text{m}$ technology node. The SSRC doping profile can be achieved by using a heavy ion with shallow projected range of distribution in the silicon crystal. Indium has properties that make it a good candidate to form SSRC doping profile as has been shown in [3.18] [3.19] [3.20]. More discussion on SSRC is given in section 2.2.3

There are three ion implantation models (Monte Carlo, Dual Pearson and Gaussian) which are incorporated into the Taurus Process and Device simulation software [3.11], to approximate the ion range distributions in the silicon substrate. This work considers the Dual-Pearson distribution model during the process simulation. However the Gaussian distribution has been used to specify an initial estimation of the projection range (R_p) and the standard deviation, σ . These two models will be discussed in section 3.2.3. No effort will be made to discuss the Monte Carlo implantation model

as it has not been used in this work due to the very long computational time scales. In the following two sections some primary concepts of stopping energy and implantation models are presented.

3.2.2 Ion stopping energy and the projection range

The physics of the ion implantation is based on the principle of the ion stopping theory, which is governed by a mechanism of energy-loss by a penetrating charged particle in a host material [3.21]. When the accelerated ions enter the substrate of the silicon crystal they start to lose the energy they acquired from the implanter. Providing a detailed analysis of ion stopping theory is beyond the objectives of this research work. However, it is helpful to discuss some basic concepts of ion stopping theory in order to understand how the ion implantation process works and how implantation models have been implemented in the process simulation software. Generally, the loss of energy by charged particles (ions) travelling through semiconductor crystal is due to collisions with electrons (inelastic collision) and nuclei (elastic collision) of the target semiconductor material. This average loss of energy per unit track length for the particle is defined as:

$$S = -\frac{dE}{dx} \quad (3.1)$$

The minus sign on the differential term signifies the loss, S and E are the total stopping power and Energy respectively. The total stopping power is approximated from the cross sections of nuclear and electron stopping powers as:

$$S = S_e + S_n = -\frac{dE_e}{dx} - \frac{dE_n}{dx} \quad (3.2)$$

The total path that the ion travels before coming to rest is known as range R , which can be calculated by integrating equation (3.2) and given by (3.3) [3:22]

$$R = \frac{1}{n} \int_0^E \frac{dE}{S_n(E) + S_e(E)} \quad (3.3)$$

Where n is an average density of electrons. Given an implantation energy E (keV), the total range R can also be estimated from the following empirical relation proposed by Mayer and Eriksson [3:21]

$$R = \left(\frac{60E}{g} \frac{M_s}{Z_s} \right) \left(\frac{M_i + M_s}{M_i} \right) \left(\frac{[Z_i^{2/3} + Z_s^{2/3}]^{1/2}}{Z_i} \right) \quad (3.4)$$

where g is the density of the host semiconductor, M_s and M_i are the atomic mass of the semiconductor material and the implanted atom, Z_s and Z_i are atomic numbers of the semiconductor material and the implanted atoms respectively. The average depth, which is the perpendicular distance from the crystal surface to the projection plane of the implanted ions, is known as the projected range R_p . It is one of the important parameters of ion implantation in semiconductor technology [3.21] [3.23]. The total path the ion travels, R , could be shorter or longer than its projected range R_p . Using the results from the empirical formula (3.4) the projection range R_p can be estimated from (3.5) and is given by

$$R_p = \frac{R}{1 + M_s/3M_i} \quad (3.5)$$

Figure 3.3 shows a 3D schematic representation of the projection range, the ion distribution and its fluctuation about R_p in an idealised semiconductor material.

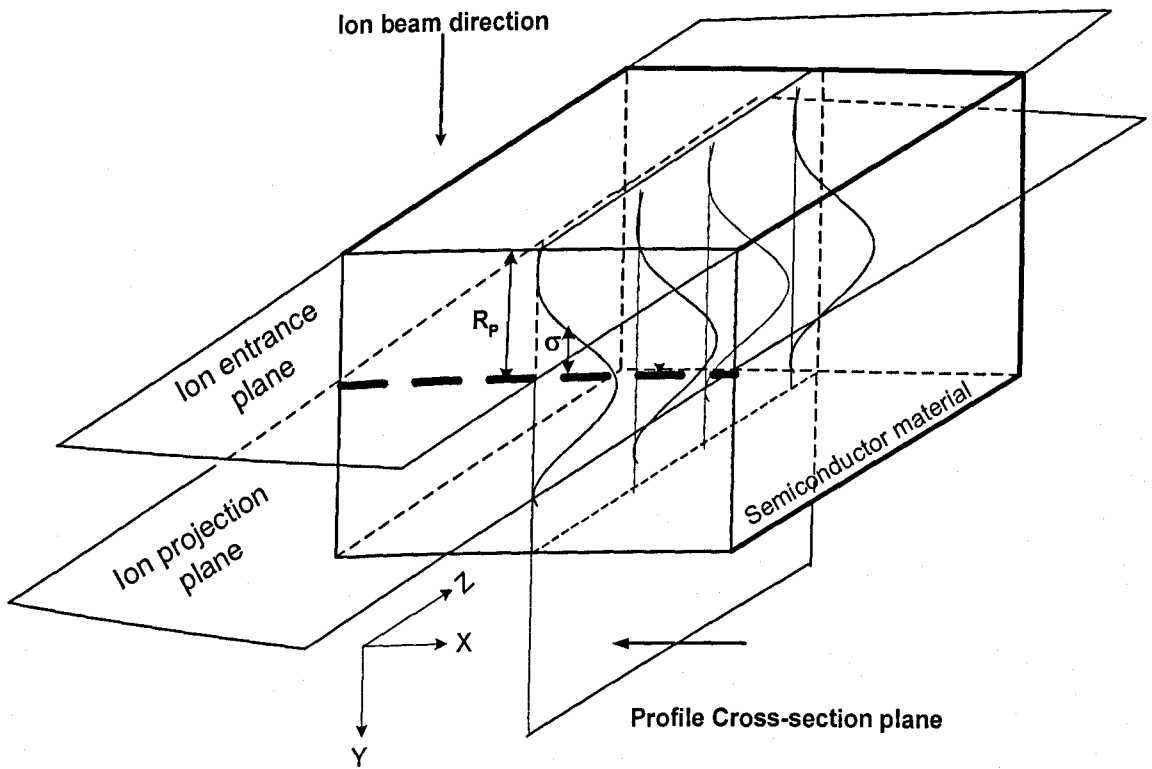


Figure 3:3 A three dimensional schematic visualisation of an average ion's projected range and its distribution in an idealised semiconductor material.

The ion beam direction for the channel doping usually perpendicular to the entrance plane. However, in some cases ion implantation may be performed with the projection angle up to 30° off the vertical axis. The projection plane is where one expects the maximum concentration of doped ions. For example in the case of our reference (Toshiba) device, the halo doping implantation is performed with tilt angle of 30° off the perpendicular to the target plane. The vertical distance between the projection plane and ion entrance plane is equivalent to the average projection distance, R_p . By rotating the cross-section plane 90° about the Y-axis to bring it in parallel to X-Y plane, the 1-D implanted ion (generally Gaussian) distribution profile could be seen clearly. Statistically speaking, the range distribution is random event and it is expected to have a profile close to a Gaussian distribution as depicted in figure 3.3 above.

3.2.3 The Gaussian and dual Pearson implantation models

The realistic channel profile in the Toshiba MOSFETs has been approximated by incorporating two virtual implantation stages in the Taurus process simulations (described in chapter 4), from which the average depth of the implanted ions and their distribution about the vertical depth have been extracted. Generally, according to the LSS theory of the projected range [3.22], the ion distribution within semiconductors and in amorphous materials is assumed to be a near Gaussian distribution given by equation (3.6) below. This is due to the statistical nature of ion stopping and scattering mechanisms.

$$N(x) = \frac{\Phi}{\sqrt{2\pi}\sigma_p} \exp\left[-\frac{1}{2}\left(\frac{(x-R_p)}{\sigma_p}\right)^2\right] \quad (3.6)$$

Here Φ is the ion flux, generally known as total ion dose of implantation per unit area [ions/cm²]. It is the total number of ions per unit area enclosed by the Gaussian curve on the projection plane.

If the maximum doping concentration is denoted by N_o [ions/cm³], it can be directly obtained from equation (3.6) for $x = R_p$.

$$N_o = \frac{\Phi}{\sqrt{2\pi}\sigma_p} \equiv \frac{0.4\Phi}{\sigma_p} \quad (3.7)$$

The Gaussian distribution has been used for preliminary estimations of the two statistical moments of ion distribution, namely the ion projected range and the standard deviation about the mean. However, the Gaussian distribution is not a particularly good model to approximate the ion implantation profile in sub-50 nanometre devices. Since the main focus of this work deals with the scaling of nano-MOSFET devices right up to the end of semiconductor technology roadmap's prediction, it is a vital to implement the correct implantation model that describes the doping profile of such devices more accurately.

The Dual Person statistical model (courtesy of the applied mathematician and statistician Karl Pearson) is a better description of the ion distribution as it includes the ion channelling effect in the silicon lattice in contrast to the simpler Gaussian distribution model, which is appropriate for another targets. Unlike the Gaussian distribution, however, the Person IV family of distributions require the third and fourth moments of the distribution in addition to the first and second moments. The first of these extra moments is the skewness of the distribution, which indicates the degree of departure from symmetry in the distribution. The second moment is kurtosis, which is the measure of how sharp the peak (*leptokurtic*) is, or alternatively how flat (*platykurtic*) the top of the distribution is. The measures of skewness and kurtosis have often been taken relative to the normal distribution. In this work, the shape of the channel profile is better described as *leptokurtic* rather than *platykurtic* (figure 4:6 in chapter 4).

Since the Pearson distribution family is a better approximation to the doping distribution compared to the simple Gaussian mode, the Dual Person implantation model has been adopted for device process simulation in this work. Therefore we, will revisit some of the basic properties and equations of the dual Person statistical distribution. The Person distribution is given by the solution of the following differential equation: [3.24, 3.25, 3.26]

$$\frac{df(x)}{dx} = \frac{(x-a)f(x)}{b_0 + b_1x + b_2x^2} \quad (3.8)$$

Where the coefficients (a , b_0 , b_1 , and b_2) of the quadratic terms are constants that can be described in terms of the four moments calculated from an arbitrary experimentally obtained distribution $f(x)$. The detailed solution and analysis of the differential equation of $f(x)$ (3.8) is given in references [3.24]. The first moment about the origin is the mean value (which is equivalent to the average projection range R_p), is given by (3.9)

$$R_p = m_1 = \frac{1}{\Phi} \int_{-\infty}^{\infty} xf(x)dx \quad (3.9)$$

The peak value of the ion concentration (N_o) is at the position determined by the projected range. The higher moments (m_2 , m_3 , and m_4) are calculated about the first moment using equation (3.9). For most impurities these parameters have been tabulated as a function of implantation energies. The major early works on implantation by Gibbons *et al* [23] were based on a Gaussian distribution and applied to amorphous materials; it is not suitable for use in the implantation model adopted in this work.

Other work on ion projection statistics calculation was reported by Peterson and Fichtner [27]. However, this work offers limited data for the indium impurity distribution properties within the silicon crystal. Therefore, it is necessary to estimate the Pearson distribution parameters for indium from the empirically calibrated equation which will be discussed in the next section. For this purpose the set of equations (3.10-3.11c) were used.

The general expression of the k^{th} moment of the distribution about the average projection range, R_p , is given by

$$m_k = \frac{1}{\Phi} \int_{-\infty}^{\infty} (x - R_p)^k f(x)dx \quad (3.10)$$

From which the standard deviation, Skewness, and Kurtosis of the distribution may be determined.

$$\text{Standard deviation} \quad \sigma_p = \sqrt{m_2} \quad (3.11a)$$

$$\text{Skew} \quad \gamma = \frac{m_3}{\sigma^3} \quad (3.11b)$$

$$\text{Kurtosis} \quad \beta = \frac{m_4}{\sigma^4} \quad (3.11c)$$

The constants a , b_0 , b_1 , and b_2 in equation (3.8) are, therefore, given in terms of the distribution moments described in (3.11a, b, c) as follows.

$$a = -\frac{m_2(4\beta - 3\gamma)}{10\beta - 12\gamma^2 - 18} \quad (3.12a)$$

$$a_0 = b = -\frac{\sqrt{m_2}\gamma(\beta - 3)}{10\beta - 12\gamma^2 - 18} \quad (3.12b)$$

$$c = -\frac{(2\beta - 3\gamma^2 - 6)}{10\beta - 12\gamma^2 - 18} \quad (3.12c)$$

3.2.4 Diffusion

Diffusion, from a device's fabrication point of view, accompanies every high temperature process step. The initial application of diffusion was for introducing dopants into a semiconductor material, changing the conductivity of a selected substrate region of interest; the formation of source and drain region of MOSFETs prior to the advent of implantation. The other process step that involves diffusion is the oxidation of the silicon surface, leading to the formation of the Si/SiO₂ interface [3.28]. This is usually accompanied by segregation which affects the doping profile of the interface. The diffusion process is also the responsible for doping redistribution during the post implantation rapid thermal annealing needed for the electrical activation of implanted ions [3.29].

As the energy of implantation and dose are vital parameters in the ion implantation process, the diffusion time and the temperature are the two major parameters governing the diffusion process [3.30]. Implantation of ions is usually done at room temperature. By contrast, the diffusion of ions needs a high temperature in the region of 800°-1200° centigrade [3.31] for silicon. In this work diffusion is mainly associated with the annealing process step. For example, in the formation of very narrow source and drain junction depths ($j_x \leq 20 \text{ nm}$), spike rapid thermal annealing (RTA) and damage recovery annealing after high energy (100-240keV) well and channel implantation were applied. In the final part of this section a brief overview of the basic equations and major types of diffusion are presented. For further reading and extensive discussions see the recommended references [3.29-3.30].

The diffusion equation used by the Taurus process simulator is based on Fick's law of solid-state diffusion given by

$$F = -D \frac{\partial N}{\partial x} \quad (3.13)$$

Where F is the diffusion flux defined as the rate of ion transfer per unit area, D and N are diffusivity [cm^2/sec] and ion concentration [atoms/cm^3] respectively. The minus sign signifies that the direction of the flux is opposite to the direction of the gradient. Rearranging the terms, and considering the fact that the rate of change in concentration with respect to time and the decrease in flux with respect to depth of diffusion are equal ($\frac{\partial N}{\partial t} = -\frac{\partial F}{\partial x}$), equation (3.13) can be generalised to give Fick's second law of diffusion given by:

$$\frac{\partial N(x,t)}{\partial t} = D \frac{\partial^2 N(x,t)}{\partial x^2} \quad (3.14)$$

In order to solve equation (3.14) various diffusion conditions have been considered. For example, concentration dependant diffusivity, constant diffusivity and temperature dependant diffusivity have all been studied [3.29].

In most cases of the present fabrication technology of CMOS devices, diffusion processes have been supplanted by ion implantation as the main method of impurity introduction to the silicon crystal. The main reasons for this were discussed above

3.2.5 Diffusion and Defects, Object-oriented Simulator (DADOS)

As CMOS devices are heading towards decananometer dimensions, variations in dopant number and their position in the channel will make device parameters vary from device to device. The threshold voltage and drive current variations are the most important parameters affected by this phenomenon. Information on the microscopically varying distribution of the dopants from realistic process simulation is vital in order to perform statistical device simulation.

In a 'standard' statistical atomistic process simulation, a continuous doping profile (obtained analytically or by using PED based process simulators) is used to generate the random discrete dopant distribution in large samples of macroscopically identical but microscopically different devices, using a variety of stochastic techniques [3.32, 3.33]. In this work we explore the possibility of using DADOS [3.7] [3.34] as a

direct source of stochastic discrete dopants distribution for statistical atomistic device simulation. DADOS is an integrated part of the Taurus Process and Device simulator. It is a 3D atomistic process simulator based on the kinetic Monte Carlo algorithm designed to study the fundamental physical diffusion phenomena in silicon. It performs atomistic level process simulation during the implantation and diffusion processes as they happen, i.e., it simulates the sequence of happenings (like point defect and emissions) at any given time and calculates Δt between individual events. That means the simulation time step is not constant [3.35].

DADOS’s integration to Taurus enables it to perform the complete atomistic process simulation steps (for example: implantation, deposition, annealing, oxidation) of realistic devices. Figure 3.4 depicts the 3D discrete doping profile of the 35 nm device, from which the dopant positions are imported from DADOS simulation.

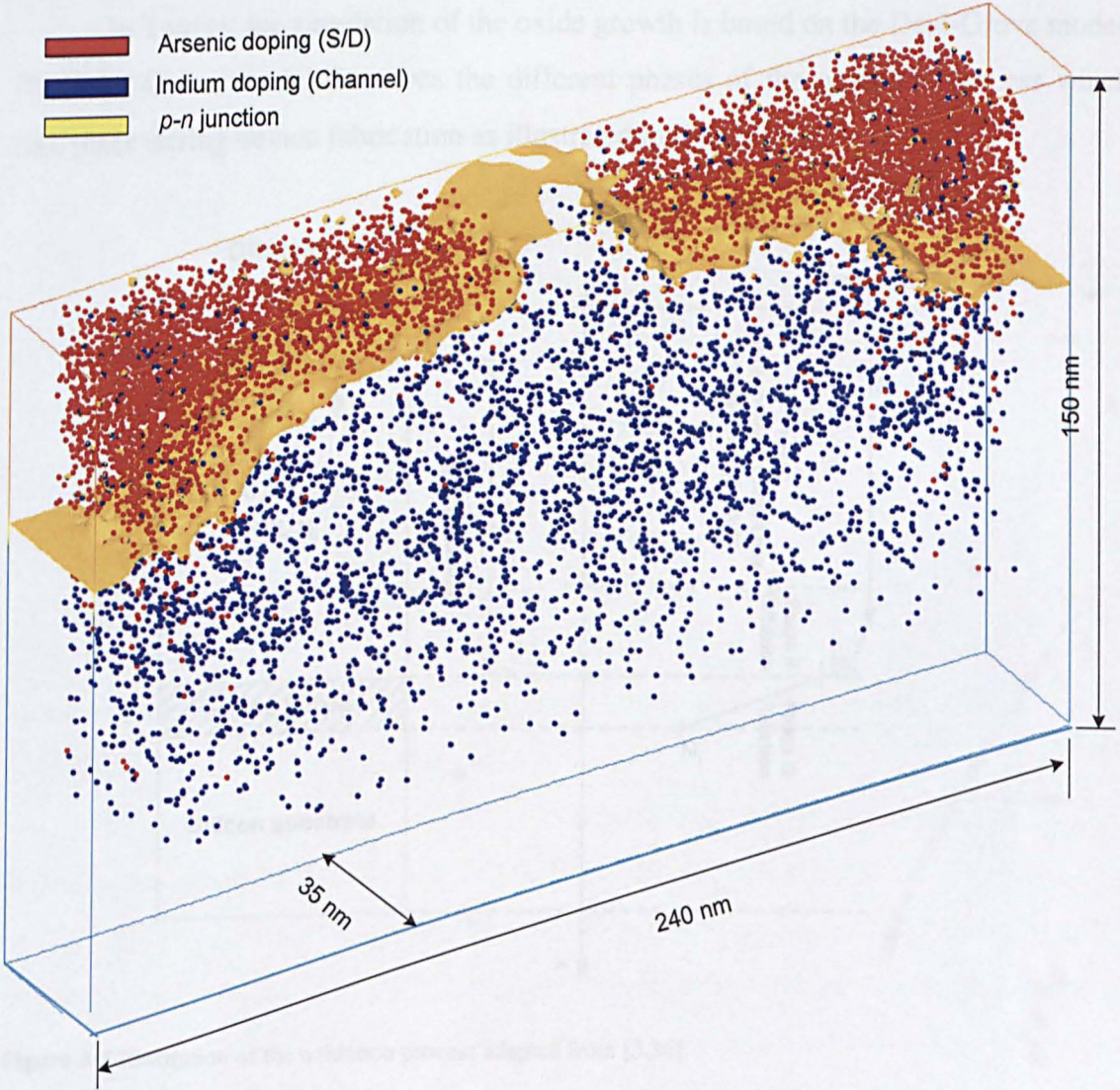


Figure 3:4 Discrete dopant distribution imported from DADOS in to the 35 nm test MOSFET. The solution domain is 240×150×35 (in nm)

3.2.6 Oxidation

The oxidation process is one of the most important steps in semiconductor fabrication resulting in grow of insulators like silicon dioxide or silicon oxy-nitride. Insulating layers are mainly used in CMOS devices for the purposes of surface protection, passivation and as a gate insulator between the gate electrode and the silicon substrate. Dielectrics may be grown using wet or dry oxidation in an oxygen rich atmosphere or by the direct deposition method on the silicon substrate (chemical-vapour deposition in the conditions of low or atmospheric pressure or plasma assisted). Figure 3.4 depicts the conventional oxidation process adopted in the standard Deal-Grove model [3.36].

In Taurus, the simulation of the oxide growth is based on the Deal-Grove model. The Deal-Grove model describes the different phases of the oxidation process which take place during device fabrication as illustrated in figure 3.5.

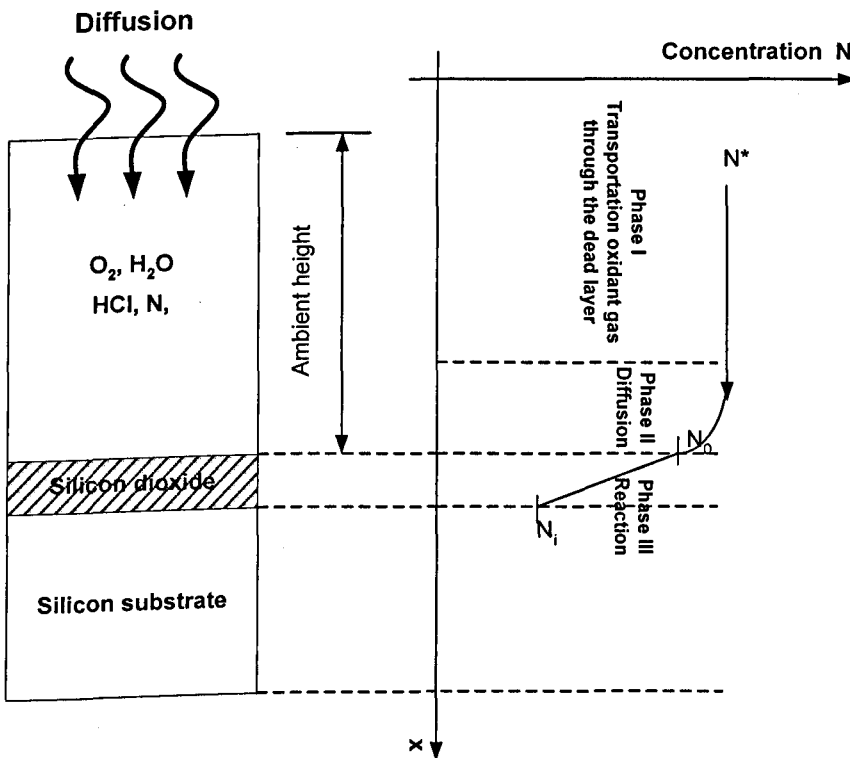


Figure 3:5 Illustration of the oxidation process adapted from [3.36]

The complete oxidation process has three main phases: the initial phase is the mass transportation of oxidant from a predefined ambient gas region to the vicinity of the silicon dioxide interface region in which the flux F_1 is given by $F_1 = h(N_g - N_0)$, where h is gas-phase transport coefficient, N_g and N_0 are the concentration of the oxidant in the ambient region and concentration of oxidant at the beginning of diffusion phase respectively. The second phase is the diffusion of the oxidant into the SiO_2 region that can be characterized by Fick's law of diffusion as

$$F_2 = D \frac{N_0 - N_i}{x_0} \quad (3:15)$$

where N_i is the oxidant concentration in the Si/SiO₂ interface. The final phase is the chemical reaction to form a homogeneous and perfect interface between the oxide and the silicon substrate. The chemical reaction that takes place on the silicon interface is given by the chemical equation, $\text{O}_2 + \text{Si} \rightarrow \text{SiO}_2$ for dry oxidation process, where O_2 is the oxidant. In the case of a wet oxidation process the oxidant is water vapour (H_2O) and the chemical equation is given by $\text{H}_2\text{O} + \text{Si} \rightarrow \text{SiO}_2 + 2\text{H}_2$. Further analysis and detailed explanations on oxidation process can be found in references [3.36, 3.37, 3.38, 3.39].

3.2.7 Gate dielectric material

SiO₂ has been instrumental in the advancement of CMOS technology in the last four decades or so. The resistance to high electric fields ($\sim 10^7 \text{V/cm}$), the near perfect interface property with silicon, and good dielectric properties are some of the key advantages of SiO₂. Despite all these qualities of SiO₂ as a gate dielectric material in CMOS devices, it may not be suitable for decanano-metre devices due to high leakage current induced by gate tunnelling through the very thin oxide and the increase in the probability of breakdown during operation.

Therefore it is very important to replace SiO₂ with high- k materials in order to maintain the pace of MOSFET miniaturization. One of such materials is the *silicon oxynitride* (SiO_xN_y) system, which has also been used by Toshiba in the fabrication of the 35nm MOSFETs used in this work. Although its relatively low permittivity (4-7) means it may not be suitable for the late stages of 65nm technology node and beyond, it has been a better replacement for the conventional SiO₂ ($\epsilon = 3.9$) in the 90 nm technology node and it is a promising candidate for 70 and 80 nm technology stages. To be

consistent with fabrication process of the realistic Toshiba device, SiO_xN_y has been adopted as the gate dielectric material rather than the conventional silicon dioxide for this project.

In the case of the Toshiba device, the SiO_xN_y gate dielectric was grown by performing an NO (for nitridation) gas annealing on very thin (1-1.2nm) base oxide surface [3.1]. 2D growth of gate oxides of 1.2nm thickness in the process simulator is computationally time consuming. Therefore, a deposition technique was used during the device calibration process.

However, by doing this one neglects segregation and the effect of stress on the process flow and eventually the likely effect of stress on the electrical characteristics of the devices during the device simulation is nullified. It should be also noticed that the permittivity of the resulting oxy-nitride system varies greatly with its stoichiometry. As the exact composition of the oxy-nitride is unknown, it may only be approximated. The approximation is given in section 4.

3.3 Device simulation

In order to understand the electrical properties of the modelled semiconductor devices one has to perform device simulations. Device simulation provides a clear understanding of general device operations, how the key parameters like threshold voltage, subthreshold slope, off and on currents behave in relation to device dimensions, and doping concentrations. It also provides microscopic information about potential carrier concentrations, field, and current distribution within the device.

In this work, 2-D and 3-D device simulations techniques have been adapted in the process of calibration, scaling and the investigation of intrinsic parameter fluctuations in MOSFETs. In this section the basic governing equations are reviewed. A brief comparison of different approaches to device simulation and concise overview of physical models used in this work are presented.

3.3.1 Equations governing device operation

In the drift diffusion approximation, operation of semiconductor devices including MOSFETs can be described by using three main equations solved self consistently, which are the building blocks of computer aided device simulation tools. One of these fundamental equations is the classical Poisson equation that describes the electrostatic

potential distribution in the devices [3.12, 3.40, 3.41, 3.42]. Despite the fact that the one dimensional Poisson equation is adequate to describe operations of the long channel devices, it may not be the case for the short and narrow channel devices. The electric field configuration in the latter case has a two/three dimensional pattern. Hence it is imperative to implement the two/three dimensional Poisson equation, which is given by:

$$\nabla \cdot (\nabla \psi) = -\frac{\rho}{\epsilon_{si}} \quad (3.16)$$

The right hand term in equation (3.15) is charge density per unit volume, which is the combination of charges contributed by electrons and holes densities and ionized donors and acceptors. Therefore, the Poisson equation can be described in terms of mobile and fixed charges as: [12]

$$\nabla \cdot (\nabla \psi) = -\frac{q}{\epsilon_{si}} [p - N_a^- + N_d^+ - n] \quad (3.17)$$

Where n and p are density of free electrons and density of free holes, N_d^+ and N_a^- are density of ionized donors and density of ionized acceptors.

The other main equation is the current continuity equations, which is split in two parts 3.18 and 3.19 for electrons and holes respectively.

$$\frac{\partial n}{\partial t} = \frac{1}{q} \nabla J_n - R_n + G_n \quad (3.18)$$

$$\frac{\partial p}{\partial t} = -\frac{1}{q} \nabla J_p - R_p + G_p \quad (3.19)$$

The terms R_n , R_p and G_n , G_p are the electron and hole recombination and generation rates respectively. The electron (J_n) and hole (J_p) are the current density terms combined from the drift and diffusion current components of the respective carriers expressed in equations 3.20 and 3.21. The remaining equations describe the current flow and are in general derived from the Boltzmann transport equation using different degrees of approximation.

$$J_n = -qn\mu_n \nabla \psi + qD_n \nabla n \quad (3.20)$$

$$J_p = -qp\mu_p \nabla \psi - qD_p \nabla p \quad (3.21)$$

Where D_n , D_p are electron and hole diffusion coefficients of electron and hole respectively. In summary, Poisson's equation describes the charge distribution in the

semiconductor devices and the continuity equation treats the transport property of carriers and the current generated by the dynamics of these carriers.

3.3.2 Comparisons of the diverse approaches to transport simulation

More complex but time consuming quantum techniques like non-equilibrium Green's function (NEGF) [3.43][3.44] and Wigner-function [3.45][3.46] are beyond the scope of this work. Therefore, this work focuses mainly on the classical simulations with quantum corrections which may be implemented to capture some of the quantum mechanical effect on sub 100 nm MOSFETs. Generally there are three main categories of device simulation methods applied to TCAD, namely the drift diffusion (DD), the hydrodynamic (HD), and the Monte Carlo (MC) device simulation approaches.

Figure 3:6 illustrates the hierarchy and the relative accuracy of the approximation of the carrier transport by the different device simulation models [3.47]. Since detailed analysis of these models is not the intention of this work, only a comparative over view of the leading concepts of the three (highlighted in figure 3:6) main approaches will be discussed.

One of the three main streams of device simulation approaches is the MC. It is a particle approach that combines periods of deterministic free-flight with stochastic scattering process implemented to approximate carrier dynamics in MOSFETs. MC uses random number generator to calculate the stochastic motion of particles and their scattering mechanisms [3.47]. The bigger the statistical sampling of the particle motion the better MC simulation approximates the distribution function.

There are different types of MC simulation methods (single particle, ensemble, and self consistent MC). For device simulations the self consistent MC approach is suitable, since it solves the Poisson equation self consistently to determine electrostatic potential distribution in the device. MC simulation does not make any assumption about the distribution function to approximate the carrier transport [3.47] [3.48] and the carriers are considered as particles rather than as a fluid, which makes the MC approach more general and accurate in approximating the carrier transport than the other two approaches.

The second approach to device simulation is the HD. In addition to the charge conservation approach inherent in the DD model (explained next), the HD model incorporates conservation of momentum (p) and energy (W) [3.49]. This will give the HD approach a better approximation of the carrier energy and temperature gradient than the DD approach. However the HD model has its own limitation. The problem with HD approach is

an over-estimation of the carrier velocity, hence unphysical high drain current in the saturation region.

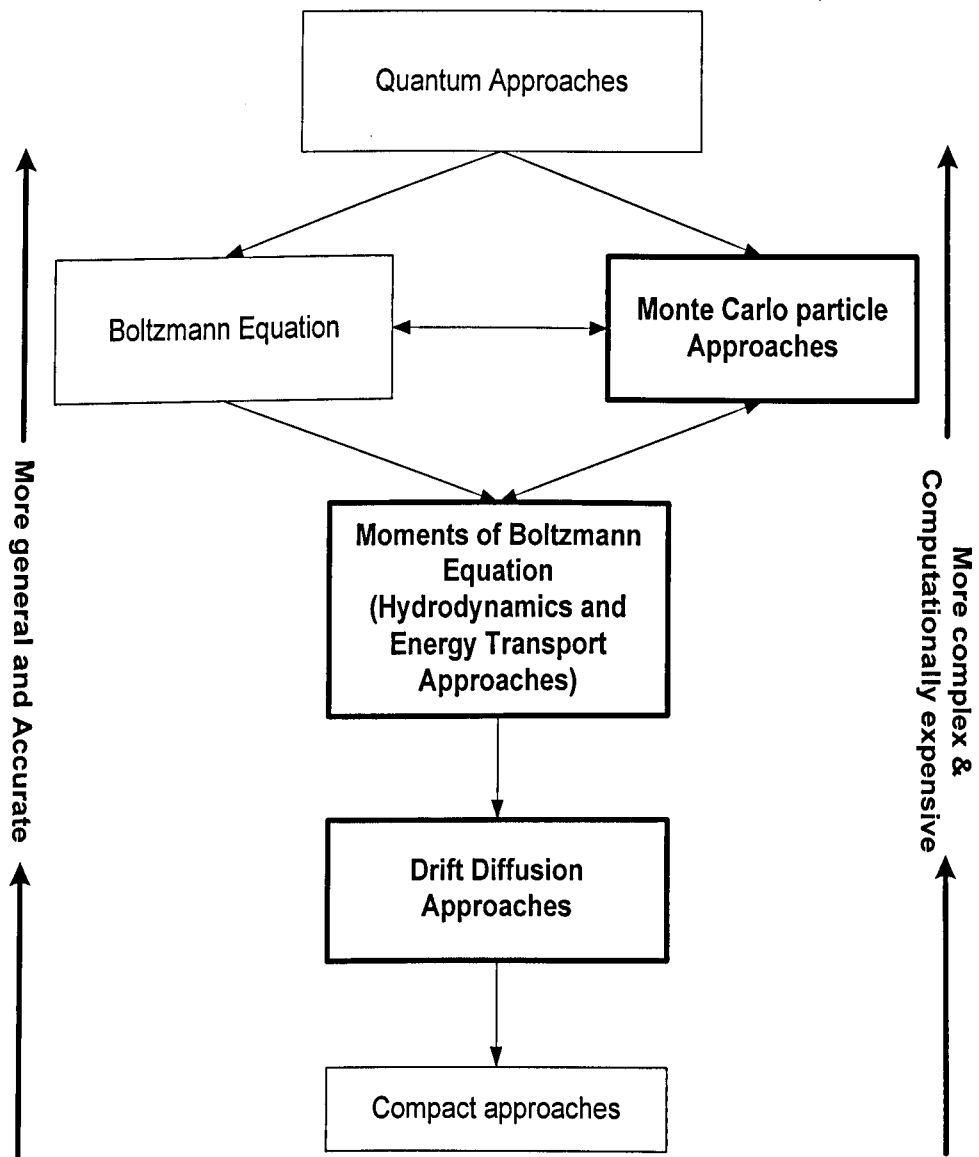


Figure 3:6 Hierarchy of the semiconductor device simulation models (after Ravaioli [3.47]).

The third approach is DD. As the name of the model suggests, DD model is solving the current continuity equation, by blending together the drift and diffusion components of current densities coupled with Poisson equation. Although it has been placed on the fourth level in the hierarchy of device simulation models as shown in figure 3:6, DD is by far the oldest and most used model for device simulations compared to the other two approaches. The main difference between the DD and the other two device simulation approaches, which are highlighted in figure 3:6, is the manner they solve the Boltzmann transport equation (BTE) in order to approximate the carrier transport and kinetic energy densities. The generalized (implicit) form of BTE is given in [3.50] as:

$$\frac{\partial F}{\partial t} + u \cdot \nabla_r f + \mathbf{F} \cdot \nabla_p F = \left. \frac{\partial f}{\partial t} \right|_c + s(x, p, t) \quad (3.22)$$

where r is the particle position, p is the momentum, \mathbf{F} is the electric field and the right hand side of the equation describes all the scattering and collision events. The solution of BTE is the distribution function (f) with seven dimensional spaces. [3.51]

The DD approach is based on the assumptions that electrons are in thermal equilibrium with lattice temperature, therefore the continuity equation is based on the conservation of charges given by

$$\partial n / \partial t = \nabla \cdot (nv) \quad (3.23)$$

but at high electric field-region carriers gain more kinetic energy influenced by the strong field. This affects the average carrier temperature which is directly proportional to the thermal kinetic energy ($\frac{3}{2} T k_B$). As a result of an increase in the average total energy ($\frac{3}{2} k_B T + \frac{1}{2} m v^2$), there will be a high carrier temperature and velocity overshoot.

As reported in [3.52] it is possible that the carrier temperature could reach up to twice the lattice temperature. Generally speaking the DD model does not estimate the carrier velocity properly [3.53], which results in drain current underestimation. The other assumption made by the DD model is to neglect the external forces during the calculation of carrier energy. This means the carrier energy is approximated solely on the basis of the local electric field [3.51] [3.54] but it depends on the global transport state of the inversion layer [3.47].

3.3.3 Implemented physical models

The device structure that will be used as a prototype model for device simulation is process simulated based on the continuous doping process. The introduction of discrete doping and atomistic aspect of device simulation will be discussed elsewhere in the thesis. Information of the device structure and doping profiles were fed into the device simulation models as an input. The drift diffusion approach has been used predominantly in the project as a compromise between accuracy and the computational demand of the 3D statistical simulations of intrinsic parameter fluctuations effects. The perpendicular electric field, the parallel electric field, and the concentration dependant mobility models were used. We have also used a modified local density approximation

(MLDA) model to approximate the quantum mechanical effect in the simulated deep sub-micron devices. The brief review on each of these models has been presented.

3.3.2.1 Concentration dependence

The concentration dependence of the carrier mobility is described by Caughey-Thomas's [3.55] empirical formula given by

$$\mu_{n,p}(x, y) = \frac{\mu_{\max} - \mu_{\min}}{1 + \left(\frac{N(x, y)}{N_{\text{ref}}} \right)^{\alpha_{n,p}}} + \mu_{\min} \quad (3.24)$$

Where, μ_{\max} and μ_{\min} are the maximum and minimum mobility respectively. N_{ref} is the reference impurity concentration used in the analytic mobility model. With the exception of $N(x, y)$, which depends on the local carrier doping concentration, all other parameters are estimated empirically by fitting to a measured data. Using default values in calibration process without modifications has been found to be a difficult job. In particular getting the right output drain current in the saturation regime was very challenging. In some cases a slight modification to the default constants given in table 3.1, which are related to the carrier mobility, was made in order to better approximate the drive current.

One reasonable explanation for the problem which occurred when the default fitting values are used to calibrate electrical characteristic of the device structure obtained from process simulation with the experimental data, is that there might be a mobility enhancement during the fabrication process of the Toshiba 35nm device. It is reported in [3.56] that a stress effects from side well, optimization of channel doping profile, and mechanical stress from the top gate layers could enhance the carrier mobility of the devices. Furthermore, it is important to note that all these parameters are fitting parameters calculated from the experimental data.

	μ_{\min} [cm ² /Vs]	μ_{\max} [cm ² /Vs]	N_{ref} [cm ³]	α_n
Taurus device (Default)	55.24	1429.23	1.072×10^{17}	0.73
Caughey-Thomas	65	1330	8.5×10^{16}	0.72
This work	60.24	1429.23	1.072×10^{17}	0.65

Table 3:1 Fitting parameters of concentration dependant mobility model

For the sake of clarity, default values for Taurus device simulator, and the original values estimated numerically in [3.55] are given in table 3.1. With the exception of α_n , all the default fitting parameters recommended in the Taurus device simulator have been maintained as they are specified in table 3.1. During the device simulation of the 35nm Toshiba device, we have changed α_n for n-type MOSFET, from its default value of 0.73 to 0.65 in order to calibrate the saturation drain current with the experimental data.

Nevertheless, to be on the safe side it is useful to test the effect of α_n on the qualitative behaviour of the universal mobility curve by plotting the empirical expression (3.24) for various values of α_n against the experimental data [3.57]. Figure (3.8) shows a plot of doping concentration against the carrier mobility for different values of the α_n in comparisons with the experimental and mobility data from Medici device simulator. As seen from figure 3.7, by changing the value of α_n between 0.65-0.73 will not alter the qualitative property of the universal mobility curve. Therefore the fitting value of α_n can safely be optimized within the reasonable range given above with out distorting the behaviour and the fitting accuracy at the model.

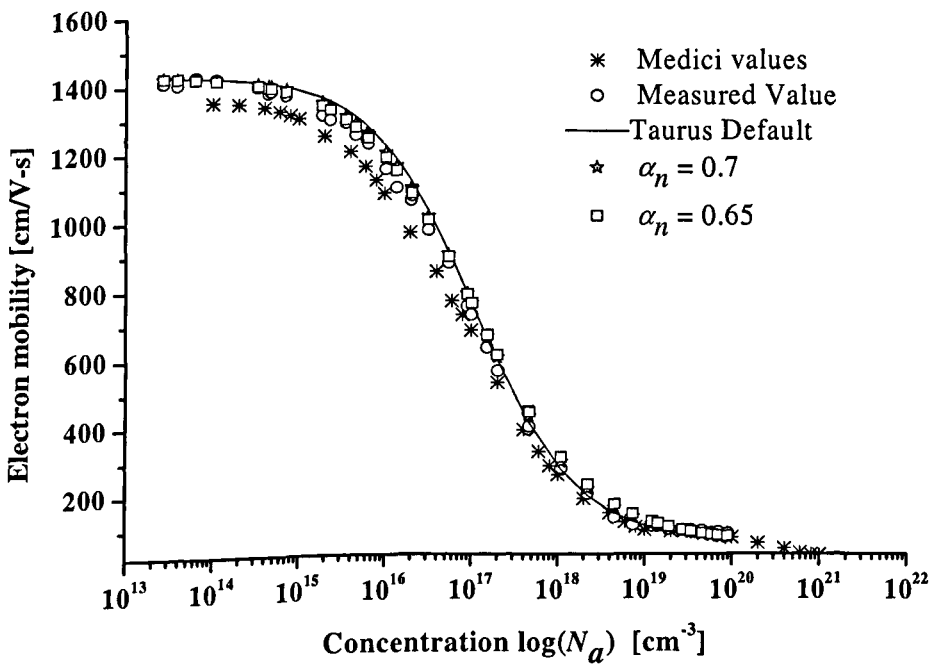


Figure 3:7 Concentration dependant mobility model with various values of α_n comparing with the default values given in the Taurus device simulator. The star and open square represents calculated values of mobility using α_n values of 0.7 and 0.65 in equation (3.15) and the continues line denote the mobility values as a function of carrier concentration specified in Medici device simulator.

3.3.2.2 Perpendicular Electric-Field Dependence

The perpendicular electric-field dependence of the carrier mobility accounts for the mobility reduction due to interface roughness scattering. The model used is based on the works of Yamaguchi [3.58]. The original model was proposed to treat the mobility as function of both parallel and perpendicular fields which is given by

$$\mu(N_d, E_{\parallel}, E_{\perp}) = \mu_o f(N_d, E_{\parallel}) g(E_{\perp}) \quad (3.25)$$

In the Taurus device simulator, equation (3.25) has been modified as shown in (3.26) so that it will take in to account only the perpendicular component of the field.

$$\mu(E_{\perp}) = \mu_o g(E_{\perp}) \quad (3.26)$$

Where $g = (1 + \alpha E_{\perp})^{-0.5}$ accounts for the property of the surface mobility and α is a fitting parameter determined from the experimental data and is estimated as $1.539 \times 10^{-5} \text{ cm/V}$ for n-channel device and $5.35 \times 10^{-5} \text{ cm/V}$ for p-channel devices. E_{\parallel} and E_{\perp} are the parallel and perpendicular electric fields respectively.

3.3.2.3 Parallel Electric-Field Dependant Mobility Model:

The parallel electric field dependence of the mobility model is modified using the expression reported in [3.55] given by:

$$\mu = \frac{\mu_0}{\sqrt[\beta]{1 + \left(\frac{E}{E_c}\right)^\beta}} \quad (3.27)$$

Where E and E_c are the electric field and critical electric field and $\mu_0 = v_{sat} / E_c$ is low field mobility for constant concentration. In Taurus Device simulation software equation (3.27) is modified to accommodate saturation velocity and the parallel electric field component in a modified form given as

$$\mu = \frac{\mu_s}{\sqrt[\beta]{1 + \left(\frac{\mu_s E_{\parallel}}{v_{sat}}\right)^\beta}} \quad (3.27b)$$

Since the doping concentration is not constant μ_0 is replaced by μ_s , which may take into account the carrier scattering effects. The constant $\beta=1$ for holes and $\beta=2$ for electrons are determined by fitting to experimental data [3.55].

3.3.2.4 Quantum Mechanical Effect - Modified Local Density Gradient Approximation

Quantum mechanical (QM) effect has a big influence on the operation of nano scale CMOS device parameters and their electrical characteristics. The miniaturization of electronic devices to a physical gate length of 7nm at the end of current ITRS roadmap demands extraordinarily thin gate oxide (0.5 – 0.6nm) and very high channel doping (well above the solid-solubility limit of known dopants in silicon). This creates a strong electric field ($\frac{\partial E}{\partial x} \approx \frac{qN_a}{\epsilon_{si}} ; E \propto \frac{1}{t_{ox}}$) responsible for the quantization of carrier motion in the direction normal to the interface [3.59]. This quantization of carriers influences the device behaviour by, for example, increasing the subthreshold voltage and decreasing the drive current of MOSFETs. To account for the QM effect during the calibration and simulation of scaled devices, the modified local density gradient approximation (MLDA) [3.60] has been included.

3.3.2.5 Glasgow Atomistic Device Simulator: 3-D

In order to investigate the intrinsic parameter fluctuations due to the discreteness and randomness of dopants position and their number, the Glasgow atomistic device simulator has been employed [3.9][3.33][3.61]. The 3-D Glasgow atomistic simulator is based on the drift diffusion approach to solve the semiconductor equations (Poisson and current continuity) with a quantum correction (density gradient).

To perform 3-D atomistic device simulation, the discrete random doping distribution of devices must be generated. It can be generated from the continuous doping profile by placing the dopant atom on the node and then deciding (by a rejection technique) whether the placed atom is a silicon or a dopant by using the probability that a dopant exists at a particular node ($p = N_a V$). Alternatively, the discrete doping profile can be extracted from the output device structure obtained by DADOS process simulator (for further explanation see section 3.25).

Furthermore, the 3D Glasgow atomistic device simulator is fairly generic. It can be adapted to perform atomistic device simulation on device structures different than the conventional MOSFET ones. For example, the silicon on insulator (SOI) [3.62] and

double gate devices have been investigated for intrinsic parameter fluctuations [3.63]. Recently, the density gradient transport model has been incorporated to the 3-D atomistic device simulator so that it is possible to investigate the quantum mechanical effect in the presence of discrete random dopants in decanano devices [3.64]. Practically the major causes intrinsic parameter fluctuation, for example LER, random dopants, and oxide thickness fluctuations can be simulated using the 3-D Glasgow atomistic device simulator.

3.4 Chapter summary

In this chapter, the main simulation methodologies and TCAD tools have been described. This includes the explanation of the fundamental concept of an integrated process and device simulation in 3D. It has been illustrated in detail how different TCAD simulation tools are systematically used during the process of this research work. By doing so, we have introduced a consistent process and device simulation approach for investigating the intrinsic parameter fluctuation due to the random dopants position and number in scaled devices according to the technology roadmap.

Comparative descriptions of the mainstream approaches of solving the semiconductor equations are presented. While the particle based Monte Carlo device simulation approach captures more accurately the high-field carrier transport the hydrodynamic model over-estimates the drive current. Although drift-diffusion is most suitable for the long channel devices, with the introduction of quantum correction it can capture reasonably well both the electrostatic behaviour and the properties of carrier in the saturated region.

When the dimensions of semiconductor devices decrease the electric field distribution increasingly shows two/three dimensional characteristics. This is due to the inhomogeneous property of the parallel electric field in the depletion region. Therefore the consideration of both the parallel electric field and the perpendicular electric field mobility model is important. In addition, to account for the spatial variations of carrier concentration, the concentration dependent mobility model has been used. Finally quantum mechanical effects have also to be taken into account by including the local density gradient approximation model. In the next chapter, the implementations of all the methodologies, which have been discussed here, are presented.

3.5 Chapter 3 references

- [3.1] S. Inaba *et al.*, "High Performance 35nm Gate Length CMOS with NO Oxynitride Gate Dielectric and Ni Salicide", *IEDM Technical Digest*, pp. 641, 2001
- [3.2] Semiconductor Industry Association, *International Technology Roadmap for Semiconductor Industry: ITRS'03 edition*, Austin, Texas: 2003
- [3.3] G. E. Moore, "Cramming more components onto integrated circuit", *Electronics*, April 19, 1965
- [3.4] F.M. Schellenberg, "Lithography: The integration of TCAD and EDA" *Semiconductor International*, Vol.27 (2), February 2004
- [3.5] M. E. Law, "Process modelling for future technologies", *IBM J. RES. & DEV*, Vol. 46, NO. 2/3, March/May 2002
- [3.6] A. Asenov, M. Jaraiz, S. Roy, G. Roy, F. Adamu-Lema, A. R. Brown, V. Moroz, and R. Gafiteanu, "Integrated Atomistic Process and Device Simulation of Decananometer MOSFETs", *SISPAD 2001*, pp.
- [3.7] M. Jaraiz, P. Castrillo, R. Pinacho, I. Martin-Bragado, and J. Barbolla, "Atomistic Front-End Process Modelling: A Powerful Tool for Deep-Submicron Device Fabrication," *Proc. SISPAD 2001*, pp. 10-17, 2001.
- [3.8] N. Strecker, V. Moroz, and M. Jaraiz, "Introducing Monte Carlo Diffusion Simulation into TCAD Tools," *Proc. ICCN 2002*, pp. 247-250, 2002.)
- [3.9] A. Asenov, "Random dopant Induced Threshold Voltage Lowering and Fluctuations in Sub-0.1 μm MOSFET's: A 3-D "Atomistic" Simulation study", *IEEE Tras. Electron Devices*. Vol. 45(12), pp. 2505-2513, 1998
- [3.10] N. E. B. Cowen, M. Jaraiz, F. Cristiano, A. Claverie, and G. Mannino, "Fundamental Diffusion Issues for Deep-Submicron Device Processing", *IEDM Tech. Dig.* pp. 333-336, 1999
- [3.11] Synopsys Taurus-TSUPREM4, Taurus-Process Reference manual, Version V2003-12, December 2003
- [3.12] Y. Taur, T. H. Ning, "Fundamentals of modern VLSI devices ", Cambridge University Press, Cambridge, UK, 1998
- [3.13] M.D.Giles, "Ion implantation", in *VLSI Technology*, edited by S. M. Sze, 2nd Edition, McGraw-Hill International Edition, pp. 327-375, 1988
- [3.14] W. Shockley, "Forming Semi-conductive Devices by Ionic Bombardment", US Patent number 2787564

-
- [3.15] B. R. Fair, "History of Some Early Developments in on-Implantation Technology Leading to Silicon Transistor Manufacturing", *Proceedings of the IEEE*, VOL. 86, NO. 1, pp. 111-137, January 1998
- [3.16] S. Tian, M. F. Morris, S. J. Morris, B. Obradovic, G. Wang, Al F. Tasch, and C. M. Snell, "A detail Physical Model For Ion Implant Induced Damage in Silicon", *IEEE Tran. Elec. Dev.* Vol. 45, pp. 1226 June 1998
- [3.17] K. W. Böer, "Survey of Semiconductor Physics" Volume 2, Barriers, Junctions, Surfaces, and Devices, Van Nostrand Reinhold, New York, 1992
- [3.18] P. Bouillon, *et al.*, "Re-Examination of Indium implantation for a low power 0.1 μ m technology", *IEDM95 Tech. Dig.*, pp. 897-900, 1995
- [3.19] P. B. Griffin, M. Cao, P. Vande Voorde, Y. -L. Chang, and W. M. Greene, "Indium transient enhanced diffusion", *Applied Physics Letters*, Vol. 73, pp. 2986, November 1998
- [3.20] H. Liao, D. S. Ang and C. H. Ling, "A Comprehensive Study of Indium Implantation-Induced Damage in Deep Sub-micrometer n-MOSFET: Device Characterization and Damage Assessment", *IEEE Trans. Elec. Dev.* Vol. 49, pp. 2254, December 2002.
- [3.21] J. W. Mayer, L. Eriksson, and J. A. Davies, "Ion Implantation in Semiconductors: Silicon and Germanium", Academic Press, New York, 1970
- [3.22] J. Lindhard, M. Scharff, and H. E. Schiøit, "*Range Concepts and Heavy ion Ranges*", *Notes on Atomic Collisions*, II Mat.-Fys Med. Dan Vid Selsk., Bind 33, No. 14, 1963
- [3.23] J. F. Gibbons, W. S. Johnson, S. W. Mylroie, "Projected Range Statistics, Semiconductors and Related Materials", Dowden, Hutchinson & Ross, 2nd Edition, 1975
- [3.24] A. Stuart, J. K. Ord, Kendall's Advanced Theory of Statistics, Volume 1, Distribution Theory, 2nd Edition, Edward Arnold, London 1994
- [3.25] Al. F. Tasch, H. Shin, C. Park, J. Alvis, S. Novak, "An Improved Approach to Accurately Model Shallow B and BF₂ Implants in Silicon", *J. Electrochemical Soc.* Vol. 136, No. 3, pp. 810-814, 1989
- [3.26] S. Selberherr, "Process and Device Modelling for VLSI", *Microelectron Reliab.*, Vol. 24, No. 2, pp. 225-257, 1984

-
- [3.27] W. P. Peterson, W. Fichtner, and E. H. targets, "Vectorized Monte Carlo Calculation for the Transport of Ions in Amorphous Targets", *IEEE Trans. Electron Devices*, Vol. 30, No. 9, pp. 1011-1017, 1983
 - [3.28] P. M. Fahey, P. B. Griffin, and J. D. Plummer, "Point defects and dopant diffusion in silicon", *Review of Modern Physics*, Vol. 61, No. 2, pp. 289, April 1989
 - [3.29] J. C. C. Tsai, "Diffusion", in *VLSI Technology*, 2nd Edition, Edited by S. M. Sze, McGraw-Hill International Edition, pp. 272-326, 1988
 - [3.30] A. M. Smith, in *Fundamentals of Silicon Integrated Device Technology*, Volume I, Edited by Burger and Donovan, Series in Solid State Physical Electronics, *Prentice Hall*, pp. 181-345, 1967
 - [3.31] S.M. Sze, *Semiconductor Devices: Physics and Technology*, 2nd Edition, *John Wiley and Sons*, USA, 2002
 - [3.32] D. J. Frank, Y. Taur, M. Jeong, and H.-S. P. Wong, "Monte Carlo Modelling of Threshold Variation due to Dopant Fluctuation", *Symposium on VLSI circuit Digest of technical Papers*, pp. 171-172, 1999
 - [3.33] A. Asenov, Gabriela Slavcheva, Andrew R. Brown, John H. Davis, and Subhash Saini, "Increase in the Random Dopant Induced Threshold Fluctuations and Lowering in Sub-100nm MOSFETs Due to Quantum Effects: A 3-D Density Gradient Simulation Study", *IEEE Trans. Electron Devices*, Vol. 48(4), pp. 722-729, 2001
 - [3.34] M. Jaraiz, G.H. Gilmer, J. M. Poate, and T.D. de la Rubia, "Atomistic calculation of ion implantation in Si: Point defect and transient enhanced diffusion phenomena", *Apply. Phys. Lett.* Vol. 68(3), pp. 409-411, 1996
 - [3.35] M. Jaraiz, P. Castrillo, R. Pinacho, L. Pelaz, J. Barbolla, G. H. Gilmer and S. C. Rafferty, "Atomistic Modeling of Complex Silicon Processing Scenarios " *Mat. Res. Soc. Symp. Proc.* Vol. 610 p.B11.1-7, 2001
 - [3.36] B. E. Deal and A.S. Grove, "General relationship for Thermal Oxidation of Silicon", *Journal of Applied Physics*, Vol.36 (12), pp. 3770-3778, Dec. 1965
 - [3.37] R. P. Donovan, in *Fundamentals of Silicon Integrated Device Technology*, Volume I, Edited by Burger and Donovan, Series in Solid State Physical Electronics, *Prentice Hall*, pp. 3-180, 1967
 - [3.38] B. E. Deal, "Thermal Oxidation Kinetics of Silicon in Pyrogenic H₂O and 5% HCl/H₂O Mixtures ", *Journal of Electrochemical Society*, Vol. 125 (4), pp. 576, 1978

-
- [3.39] C. P. Ho. and J.D. Plummer, "Si/SiO₂ Interface Oxidation Kinetics: A Physical Model for the Influence of High Substrate Doping Levels", *Journal of Electrochemical Society*, Vol.126(9), pp. 1516-1530, 1979
- [3.40] T. Ando, B. A. Fowler, F. Stren, "Electronic properties of two dimensional systems", *Reviews of Modern Physics*, Vol. 54, No. 2, pp. 437, April 1982
- [3.41] S. M. Sze, *Physics of Semiconductor Devices*, Wiley International science, 1981
- [3.42] S. Selberherr, "MOS Device Modelling at 77 K", *IEEE Trans. Electron Dev.* Vol.36, pp. 1464-1474, 1989
- [3.43] Z. Ren, R. Venugopal, S., Datta, and M. Lundstrom, "Examination of design and manufacturing issues in a 10 nm double gate MOSFET using non-equilibrium Green's function simulation", *IEDM Tech. Dig.*, pp, 107-110, December 2001
- [3.44] S. Datta, "The non-equilibrium Green function (NEGF) formalism: An elementary introduction", *IEDM Tech. Dig.*, pp, 703-706, December 2002
- [3.45] L. Shifren, and D. K. Ferry, "Wigner Function Based Monte Carlo Approach for Accurate Incorporation of Quantum Effects in Device Simulation", *Journal of Computational Electronics*, 1: 55-58, 2002
- [3.46] P. Bordone, and C. Jacoboni "Wigner Paths for Quantum Transport", *Journal of Computational Electronics*, 1: 67-73, 2002
- [3.47] U. Ravaioli "Hierarchy of simulation approaches for hot carrier transport in deep submicron devices", *Semicond. Sci. Technol.*, Vol. 13, pp. 1-10, 1998
- [3.48] K. Tomizawa, *Numerical Simulation of Submicron Semiconductor Devices*, Artech House, Boston, London, 1993
- [3.49] P. Lugli, "The Monte Carlo Method for Semiconductor device and Process Modelling", *IEEE Trans. On Computer-aided Design*, Vol. 9(11), 1999
- [3.50] M. Lundstrom, *Fundamentals of carrier transport*, Second edition, Cambridge, UK: Cambridge University Press, pp. 119-157 , 2000
- [3.51] S. Selberherr, *Analysis and Simulation of Semiconductor devices*, New York: Springer-Verlag, 1984
- [3.52] T. Grasser, T - W. Tang, H. Kosina, S. Selberherr, "A Review of Hydrodynamic and Energy-Transport Models for Semiconductor Device Simulation", *Proceedings of the IEEE*, Vol. 91(2), pp. 251-274, 2003
- [3.53] J.D. Bude "MOSFET Modelling Into Ballistic Regime", in *Proc. Simulation Semiconductor Processes and Devices, SISPAD 2000*, pp. 23-26, 2000

-
- [3.54] T. Grassler, H. Kosina, and S. Selberher, "Consistent Comparison of Drift-Diffusion and Hydro-Dynamic Device Simulations", in *Proc. Simulation Semiconductor Processes and Devices, SISPAD' 99*, pp.151-154, 1999
- [3.55] D.M. Caughey, and R. E. Thomas, "Carrier Motilities in Silicon Empirically Related to Doping and Field", *Proceedings of the IEEE*, Vol. 55, pp. 2192, 1967
- [3.56] T. Sugii, K. Watanabe, and S. Sugatani, "Transistor Design for 90nm-Generation and Beyond", *Fujitsu Sci. Tech. J., Special Issue: Advanced Silicon Device Technologies*, Vol. 39 (1), pp.9-22, 2003
- [3.57] W.R.Thurber, R. L. Mattis, and Y.M. Liu, and J. J. Filliben, "Resistivity-Dopant Density Relationship for Phosphorus-Doped Silicon", *J. Electrochem. Soc.* Vol. (8), 1807-1812, 1980
- [3.58] K. Yamaguchi, "Field-Dependant Mobility Model for analysis of MOSFET's", *IEEE Trans. Electron Dev.*, Vol. 26, No. 7, pp. 1068-1074, 1979
- [3.59] F. Stern, and W. E. Howard, "Properties of Semiconductor Inversion Layers in The Electric Quantum Limit", *Phys. Rev.* Vol. 163 (3), 1967
- [3.60] G. Paasch and H.Ubensee, "A Modified Local Density Approximation", *Phys. Stat. Sol. (b)*, Vol. 113, pp. 165-178, 1982
- [3.61] A. Asenov, Andrew R. John H. Davies, and Subhash Saini "Hierarchical Approach to "Atomistic" 3-D MOSFET Simulation", *IEEE Trans. Computer-Aided Design of integrated Circuit and Systems*, Vol. 18 (11), pp. 1558-1565, 1999
- [3.62] A. R. Brown, FikruAdamu-Lema, and Asen Asenov, "Intrinsic parameter fluctuations in nanometre scale thin-body SOI devices introduced by interface roughness", *Super lattices and Microstructures*, Vol. 34, pp. 283-291, 2003
- [3.63] A. R. Brown, J. R. Watling, and A. Asenov, "Intrinsic Fluctuations in Sub 10-nm Double-Gate MOSFETs Introduced by Discreteness of Charge and Matter", *IEEE Trans on Nanotechnology* , Vol. 1(4), pp. 195 – 200, 2002
- [3.64] A. Asenov, G. Slavcheva, A. R. Brown, J. H. Davies, and S. Saini, "Increase in the Random Dopant Induced Threshold Fluctuations and Lowering in Sub-100 nm MOSFETs Due to Quantum Effects: A 3-D Density-Gradient Simulation Study", *IEEE Trans. Electron Devices*, Vol. 48 (4), pp. 722-729, 2001

4

Scaling of the Conventional MOSFET Devices

The previous section outlined the integrated process and device simulation methodology that has been followed throughout this research in the modelling and analysis of real 35 nm MOSFETs. This chapter deals with the implementation of this methodology and the scaling of the 35 nm transistor to its ultimate limits. The first section presents the general concept of the calibration process and the practical steps which have to be followed to achieve both a good understanding of the device structure and good agreement with the measured device characteristics. The structure and the electrical properties of the well-calibrated 35nm CMOS device which is used as a starting point of the scaling study is presented in the second section. Section three describes work on the scaling of the MOSFETs based on predictions for the future four generations of technology nodes by the ITRS. Additionally, the device structures and the electrical and physical properties of the scaled devices have been analysed. Further discussions and conclusions are presented in the last section of this chapter

4.1 Calibration strategy and its role in device simulation

In general, calibration is a process in which the parameters used in the simulation are tuned to deliver repeatable results in all simulation conditions within the space of reasonable input parameters [4:1] [4:2].

In device simulation, the calibration process helps to validate our physical models by comparing simulated results with real device characteristics. Hence, the calibration process from the device simulation point of view can be defined as a methodical approach which leads to the synchronisation between the simulation results and valid experimental data.

This matching process is very important in order to validate the simulation results, to build confidence in the models and to allow reliable analysis based on these results. Therefore, the aim of calibration is to tune the model parameters in order to produce simulation results which are as close as possible to the experimental data. At the same time it is very important to make sure that all the input parameters that have been used and tuned in the calibration process are physically meaningful, leading to trustworthy simulation results.

Therefore, a carefully and systematically executed calibration is the stepping-stone to a successful device simulation and modelling strategy. This means that calibration plays an important role in the device simulation practice. One of the advantages of device calibrations is the possibility of implementing various physical models with different degrees of complexity. Unfortunately, all the models typically available in TCAD tools are not expected to give identical simulation results for the same devices structure. At the same time, some of them are not sufficiently robust to be applied in the simulation of nano-scale MOSFETs.

Hence, to stick to a particular model throughout the simulation practice independent of the device size or the bias conditions is highly unwise. It is always important to evaluate the available models and it is essential to perform a test in order to select an optimum simulation model both in terms of accuracy and computational efficiency

Through device calibration it is possible to achieve both accurate and physically meaningful simulation results. This in turn can empower the technologists and the device engineers to develop the optimum device structure and the corresponding fabrication steps that can deliver the desired electrical parameters. It is important to note that the default parameters in the relevant TCAD tools are not always optimal for particular types of devices and as a rule do not deliver good agreement with the experimental data straight away unless calibration has been performed systematically and thoroughly.

In order to achieve a reliable and effective simulation based device scaling, a systematic calibration methodology has been developed and applied in the course of this work, which is outlined in the next subsection.

4.1.1 Systematic calibration

It is a good strategy in the calibration process to start from the simplest physical model and work up the ladder to the more complex and comprehensive models. Although simple models are inherently attractive because of a small number of input parameters and computational efficiency, in most cases they can only provide basic qualitative and semi-quantitative results. The introduction of more complex models is usually needed to improve the quantitative agreement. Therefore a hierarchical strategy can be applied to the device calibration process without compromising its final outcome.

The calibration starts with the gathering of all the important available information about the structure and the characteristics of the calibrated device. The careful analysis of the device structure and characteristics provide indications about the appropriate models. The choice of appropriate physical models prompts the beginning of the calibration processes.

The overall calibration strategy is illustrated in figure 4:1. Detail of the important calibration steps with the corresponding models of choices, will be discussed in the next three sections. The systematic calibration methodology illustrated in figure 4:1 can be summarised in to the following 6 step algorithm.

- 1 Start from the simple operation conditions which can be reproduce by simple models, for example, low field part of the I - V characteristics which can be described with constant mobility models.
- 2 Adjust the channel doping profile and the halo-doping concentration at constant mobility to capture the device electrostatics by matching the threshold voltage V_T and the subthreshold slope- S . Test the effect of quantum corrections on V_T ,
- 3 At low drain voltage increase, the complexity of the mobility model to capture the perpendicular electric field and concentration dependence of the mobility
- 4 At high drain voltage refine the doping profile to capture the drain induced barrier lowering (DIBL) effects and introduce the lateral field dependence of the mobility to capture the velocity saturation effect
- 5 Examine the impact of non-equilibrium transport by switching-on hydrodynamic tools

- 6 Compare the simulation result with the experimental data and analyse the developing variation between the two. If errors are outside of margin of tolerance, go back to the starting point or to a relevant calibration step for the necessary input parameter adjustments.

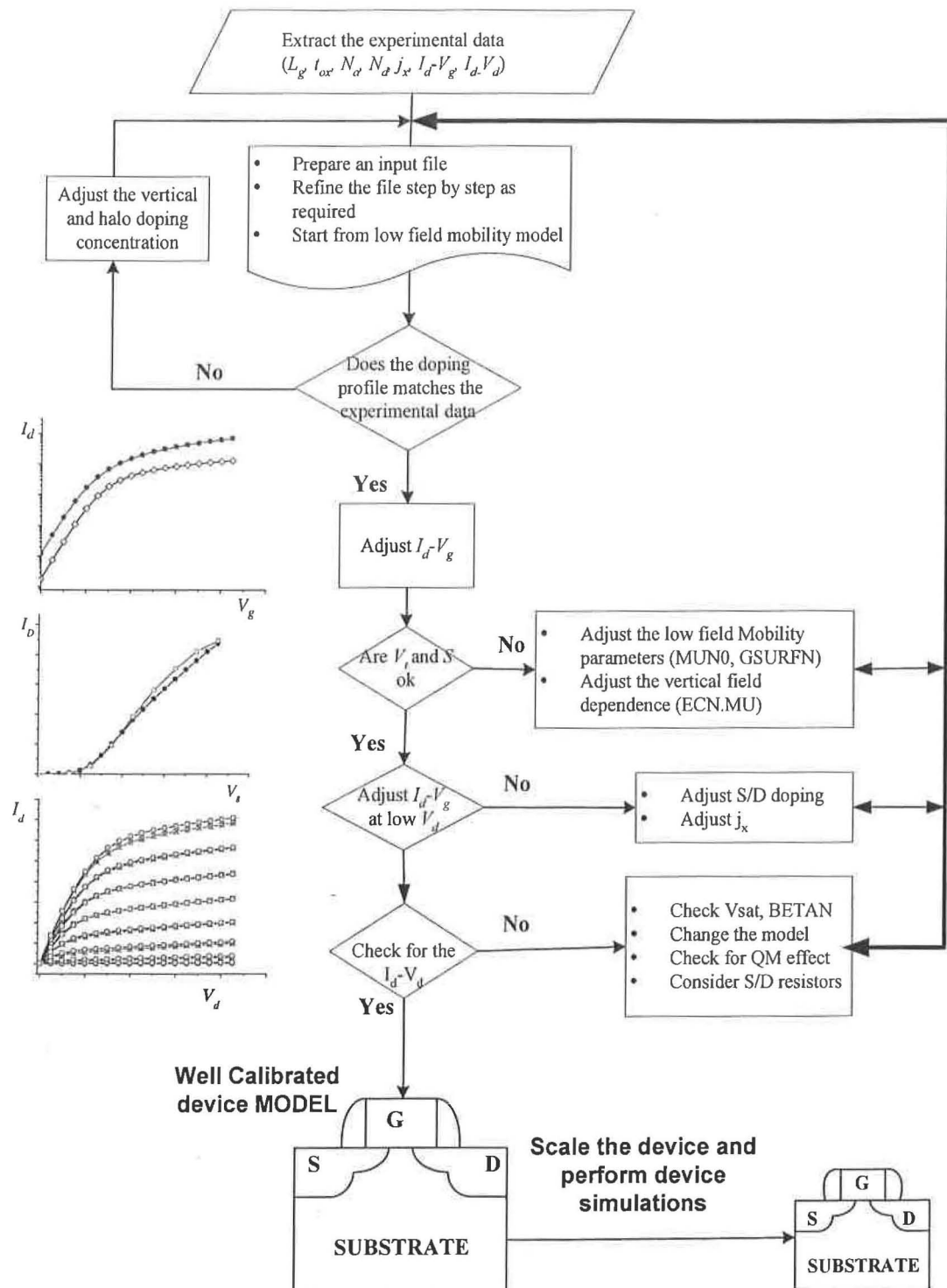


Figure 4:1 Flow chart that illustrates the systematic calibration methodology

4.2 The structure of the real 35 nm gate length MOSFET

This subsection introduces the structure and the critical design parameters extracted from the 35 nm gate length MOSFET fabricated by Toshiba which has been used as a reliable experimental source for calibration of the simulator tools applied in this project. The quality of the experimental data is crucial for the successful calibration. Some of the most important parameters for the above device presented in table 4.1 below have been carefully extracted from published data [4.3] and from private communication with the authors.

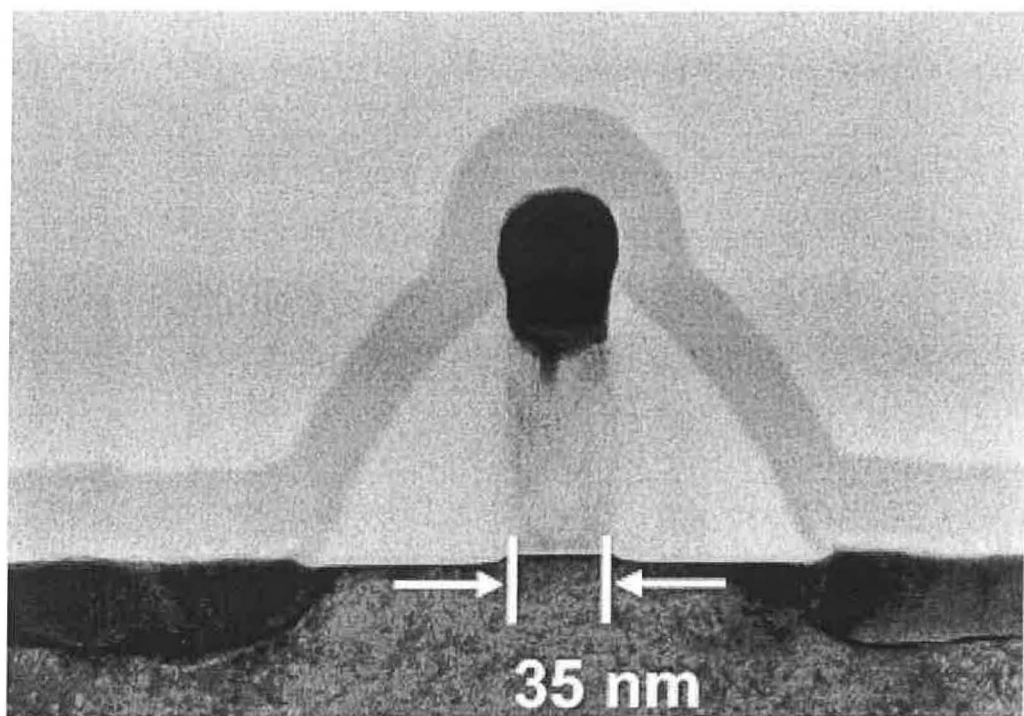


Figure 4:2 Transmission electron micrograph (TEM) photograph of the 35 nm *n*-channel device from which all data and process information have been obtained for this work. Reference [4.3]

In respect of the channel profile engineering of the 35 nm transistor, the authors of [4.3] have investigated three different cases of indium implantation in order to establish the optimum doping profile for their *n*-channel MOSFET. The first case introduces a high dose of indium implantation with no halo doping. The second case introduces intermediate channel dose and intermediate halo doping concentration. The final case introduces low dose of channel implant and high dose of BF_2 halo implantation. The implantation

conditions in the above three cases have been simulated based on the information provided by the authors of [4.3].

A significant effort has been invested in reproducing a realistic device structure using proper process simulation steps. Figure 4:2 shows the transmission electron microscopy cross-section view of the Toshiba 35 nm physical gate length MOSFET, from which the essential structural data were extracted for the calibration process in this work.

The extracted structural data are summarised in table 4:1 for both n- and p-channel devices. The corresponding electrical parameters are summarised in table 4:2. Although the 35 nm gate length transistor is not specified explicitly in the ITRS, its electrical and structural parameters are very close to that of the 37 nm physical gate length MOSFET which is required for late stage of the 90 nm technology node.

	Junction depth x_j [nm]	Oxide thickness t_{ox} [nm]	S/D doping [ion/cm ²]	Channel doping [ion/cm ²]	Halo doping [ion/cm ²]
n-MOSFET	20	1-1.2	Unpublished data *	$1.0 - 5.0 \times 10^{13}$	$< 3.0 \times 10^{13}$
p-MOSFET	33	1-1.2	Unpublished data *	$1.0 - 4.0 \times 10^{13}$	Unpublished data *

Table 4:1 Important device dimensions and doping information of the 35 nm gate length p-type MOSFET.

	Threshold voltage [mV]	Subthreshold slope [mV/dec]	Subthreshold leakage current [nA/ μ m]	Supply voltage [mV]	Saturated drive current [μ A/ μ m]
n-MOSFET	No data available	86.1	100	850	676
p-MOSFET	Not data available	92.3	100	850	272

Table 4:2 Electrical parameters of the realistic Toshiba 35 nm gate length n-type MOSFET

* Unpublished data have been obtained through personal communication with the authors.

4.2.1 The Doping profiles of the channel, and the source and drain extensions

In the course of this work, the initial target of the process simulation was to calibrate the process simulation sequence performed with Taurus Process, a commercial simulation tool, in order to reproduce source and drain extension doping profiles of the experimental device. The amount of doping and its distribution in the active region of the MOSFET is a critical factor determining the operation of device.

The short channel effects in decanano meter scale transistors are a major limitation that hinders the scaling of these devices. The main approach to suppress short channel effects is to increase the doping concentration in conjunction with optimum channel doping profile engineering. It is therefore very important for the calibration of the electrical simulation to obtain the correct channel doping profile from realistic process simulation.

Prior to proceeding with the calibration of the process simulation, it is important to study and review the detailed doping profile for 35 nm MOSFET in both the channel region and source/drain extensions. Indium (**In**) and arsenic (**As**) have been implanted to give a super retrograde (low-high-low) channel profile in both *n*-MOSFET and *p*-MOSFET respectively. For source and drain extensions formation, (**As**) for *n*-MOSFET and boron (**B**) for *p*-MOSFET were implanted.

The study of the channel **In** profile of the *n*-channel MOSFET, suggests that the channel doping profile (figure 4.4) doesn't follow closely a normal distribution typical for ion implantation and the follow-up rapid thermal annealing (RTA) steps. This complex **In** profile is difficult to reproduce in the process simulation. The main reason is that realistic models for **In** implantation and diffusion were not included in Taurus process simulator at the time of this work. The process that causes the tailing in the indium distribution on both sides of the peak during the RTA are not well understood. This might be one of the reasons why it is not incorporated in Taurus process simulator.

We have, therefore, adopted a pragmatic approach when simulating the **In** distribution in the channel. Based on the complexity of the **In** doping profile of the 35 nm device; we can make two pragmatic observations regarding the doping profile of the channel region:

1. The final **In** doping profile in the channel of the 35 nm device including the tailing of **In** profile due to transient effects during RTA in the real fabrication process could be achieved by performing two 'virtual' ion implantations. The result of these successive implantations and diffusion during the RTA delivers the **In** doping profile measured in the channel region of the transistor. Each implantation has different parameters, including dose, energy, projection range, and the standard deviation.
2. The doping profile associated with each of these 'virtual' implantations has a nearly Gaussian distribution (figure4:3). This pragmatic approach gives as an opportunity to establish two sets of Gaussian distribution parameters that may characterize the ion distributions resulting from the two 'virtual' implantation steps. The sum of these two distributions provides an initial approximation to the measured **In** doping distribution.

4.2.1.1 Initial estimation of implantation parameters

Following the pragmatic approach for modelling of the **In** distribution outlined in the previous section, the two Gaussian distribution functions, which represent the two 'virtual' implantations, are given by equations 4.1 and 4.2.

$$n_1(x) = n_{01} \exp \left[-\frac{1}{2} \left(\frac{x - R_{p1}}{\sigma_1} \right)^2 \right] \quad (4.1)$$

$$n_2(x) = n_{02} \exp \left[-\frac{1}{2} \left(\frac{x - R_{p2}}{\sigma_2} \right)^2 \right] \quad (4.2)$$

The two Gaussian distributions have different projection range R_{p1}, R_{p2} standard deviation σ_1, σ_2 and doses related to the maximum concentrations n_{01}, n_{02} . The task is to deduce the above parameters from the experimental distribution, so that the calibration process can be performed progressively.

The deduced values for projected range, standard deviation and doses can then be applied in the process simulation using for example the popular Dual-Pearson implantation model.

We expect that this approach will provide a reasonably good approximation to the final doping profile implantation parameters, reasonably close to the measured **In** distribution in the channel. The schematic illustration of the two Gaussian distributions is shown in figure 4:1. By observing the individual properties of these distributions and their amalgamated effect on the final profile after implantation and annealing, it is reasonable to suggest that the narrow Gaussian (denoted as Gaussian₁) influences the peak shape of the doping distribution in the neighbourhood of the maximum concentration located at the projection range (R_{p1}). On the other hand the wider Gaussian (denoted as Gaussian₂) is affecting the exponential tail of the doping profile, which mainly happening due to the channelling [4:4] and transient diffusion during the RTA.

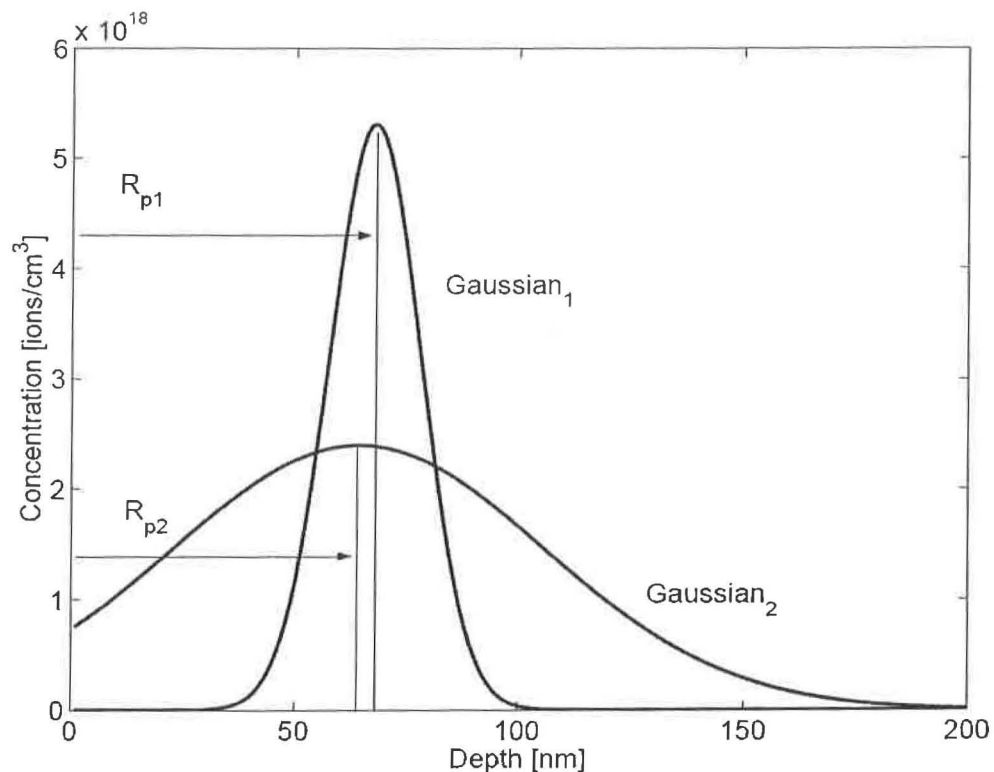


Figure 4:3 Two Gaussians are illustrated with their corresponding projection range, (R_{p1} and R_{p2}) and standard deviations (σ_1 and σ_2). These two curves represent the two arbitrary Gaussians distributions, given by equations 4.1 and 4.2. The point of this figure is to illustrate the practical approach, which highlights that the combined effect two virtual implantation may yield the final doping profile of the channel implantation.

The curve defined by the sum of the two equations in (4.1, 4.2) has been fitted to the measured experimental data presented in figure 4:4 below. As a result of this fitting, one can determine the preliminary values of the parameters of the two equations, which later will be empirically tuned to mach the measured **In** distribution after the implantation and the annealing steps during calibration. The parameter values that are given in table 4.3 correspond to the fitted curve shown in figure 4:4. These values are not the final results of the calibration process. But they give a reasonable initial distribution of the channel dopants for the calibration process, which will then include further dopant redistribution as a result of post implantation annealing and other high temperature fabrication steps.

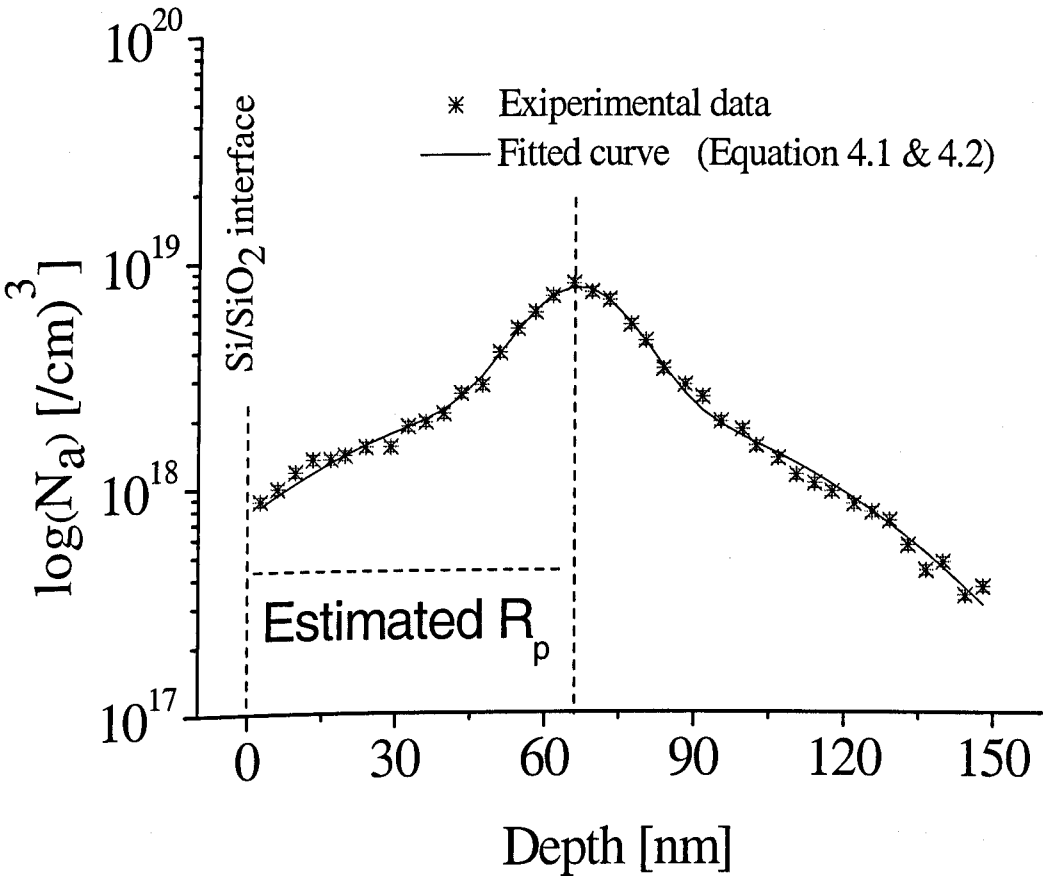


Figure 4:4 The sum of the two Gaussian equations in 4.1 and 4.2 is fitted to the Experimental data in order to approximate the combined projection range and standard deviation. The symbols represent the measured Toshiba data and the line represents the sum of the two equations. The approximated values obtained from this fitted curve are given in table 4:3

As one can notice from the figures in table 4:3 below, the error margins for standard deviations are slightly higher than the error margins of the maximum concentration and the average projected range value. On the other hand, the error margins for the average projection range and the maximum concentration are within a reasonable tolerance range. We can adjust the standard deviation as well as the other parameters of the distribution in order to match the experimental data. The process continues until optimum value of each parameter of the ion distributions. It is also reasonable to say that the calibrated results shall reflect these error tolerances.

	$R_{p1} \& R_{p2}$ [nm]	$\sigma_1 \& \sigma_2$ [nm]	Φ -Dose [ions/cm ²]	$n_{01} \& n_{02}$ [ions/cm ³]
Gaussian _1	$66.8 \pm 0.3\%$	$10.2 \pm 5\%$	1.4×10^{13}	$5.3 \times 10^{18} \pm 2.2\%$
Gaussian _2	$63.4 \pm 1.4\%$	$41.7 \pm 7.137\%$	2.3×10^{13}	$2.4 \times 10^{18} \pm 1.4\%$

Table 4:3 Initial estimation of numerical values obtained from the curve fitting of the two equations (4.1a, b) to the experimental data of the Toshiba device. The fitted curve is illustrated in figure 4.3 with the highlight of the projection range

In addition to the values given in Table 4:3 for the corresponding ion range distributions of the two virtual implantations, it is also possible to analytically approximate the resultant values of the two distributions. Using equations 4.1 and 4.2 in conjunction with parameters in table 4.3 the overall ion projected range, standard deviations, and the total ion dose have been presented in the next section.

4.2.1.2 Analytical calculation of R_p, σ_p and Φ of the two Gaussians

In order to compare extracted parameters with the experimental values it is useful to find the combined R_p, σ_p and Φ . The resultant projected range can be calculated as the

first moment of the combined two Gaussian equations integrating from the Si/SiO₂ interface ($y = 0$) to the device depth $y = 0.16 \mu\text{m}$ used in the simulation domain.

$$\begin{aligned}
 R_p &= \frac{1}{\Phi} \int_{-\infty}^{\infty} xf(x)dx \\
 &= \frac{1}{\Phi} \int_0^{160E-7} x \left[5.3 \times 10^{18} e^{\left[\frac{-1}{2} \left(\frac{x-66.8E-7}{10.2E-7} \right)^2 \right]} + 2.4 \times 10^{18} e^{\left[\frac{-1}{2} \left(\frac{x-63.4E-7}{41.7E-7} \right)^2 \right]} \right] dx \quad (4.3) \\
 &= 67.29 \text{ nm}
 \end{aligned}$$

The total dose (Φ) can be estimated as the integral of the sum of the two distributions given by equations 4.1 and 4.2

$$\begin{aligned}
 \Phi &= \int_0^{\infty} \left[5.3 \times 10^{18} e^{\left[\frac{-1}{2} \left(\frac{x-66.9E-7}{10.2E-7} \right)^2 \right]} + 2.4 \times 10^{18} e^{\left[\frac{-1}{2} \left(\frac{x-63.4E-7}{41.7E-7} \right)^2 \right]} \right] dx \quad (4.4) \\
 &\approx 3.7 \times 10^{13} [\text{ions} / \text{cm}^2]
 \end{aligned}$$

This determines the total dose of indium in the channel, which is 3.658×10^{13} ions/cm². The estimated dose of the implantation is well within the range of the implantation used in the fabrication of the Toshiba devices published in [4.3], which is in the range of $1.0 - 5.0 \times 10^{13}$ ions/cm².

The standard deviation of the combined distribution is the square root of the second moment of the ion distribution given by $\sigma_p = \sqrt{m_2}$ and can be calculated by using the analytical expression for the standard deviation as shown in equation 4.5:

$$\begin{aligned}
 \sigma_p &= \left[\frac{1}{\Phi} \int_0^{160E-7} (x - R_p)^2 \left(5.3 \times 10^{18} e^{\left[\frac{-1}{2} \left(\frac{x-66.8E-7}{10.2E-7} \right)^2 \right]} + 2.4 \times 10^{18} e^{\left[\frac{-1}{2} \left(\frac{x-63.4E-7}{41.7E-7} \right)^2 \right]} \right) dx \right]^{\frac{1}{2}} \quad (4.5) \\
 &= 34.8 \text{ nm}
 \end{aligned}$$

4.2.1.2 Calibrated channel doping profile: *n*-MOSFET

The **In** distribution in the channel obtained as a result of the complete process simulation is illustrated in figure 4:5. The analytically calculated projection range of the two combined Gaussians (see figure 4.4) and the simulated projected range of the doping profile, shown in figure 4:5 are 66.8 nm and 65.7 nm respectively. The analytically calculated and the simulated values of the projection ranges are slightly bigger by 2.77% and 1.107% respectively compared with the measured value of 65 nm. Therefore the calculation of R_p from our empirically fitted curve is in a very good agreement with the measured projected range of indium ions in the real 35 nm transistor.

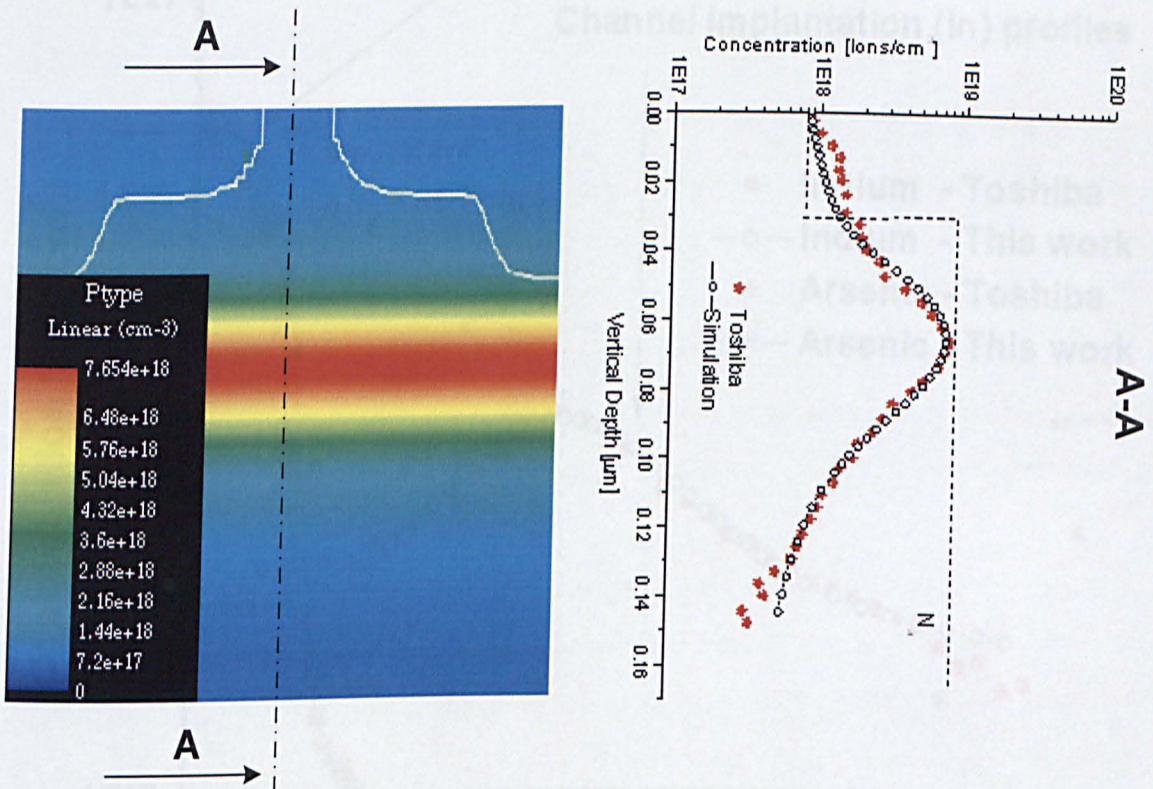


Figure 4:5 The resultant (calibrated) channel doping profile of *n*-channel MOSFET, approximated from the two Gaussian distributions given by equations 4.1 and 4.2. The figure on the right is obtained from the cross section along the vertical position at the axis denoted A-A.

During the calibration process, the projected ranges and the standard deviations, which are approximated by equations in 4.1 & 4.2, have been further adjusted in order to obtain a good agreement between the simulated and the experimental ion distribution in the channel region of the device. This is necessary because the distribution of **In** ions after the implantation evolves as a result of the diffusion associated with the RTA.

The image on the left side of figure 4:5 illustrates the 2D indium doping profile of the calibrated n-MOSFET. The maximum concentration of indium implantation in the range of $7.5\text{--}8.5 \times 10^{18} \text{ /cm}^3$ is located at the average range of implanted ions along the incident-beam direction. As it can be seen from the 1-D plot of the **In** doping profile illustrated in figure 4:6, there is an excellent agreement between the simulation and the measured data. The 1-D profile plot is obtained from the cross-section (A-A) at the middle of the channel from the simulated 2-D channel profile shown on the right side of figure 4:5.

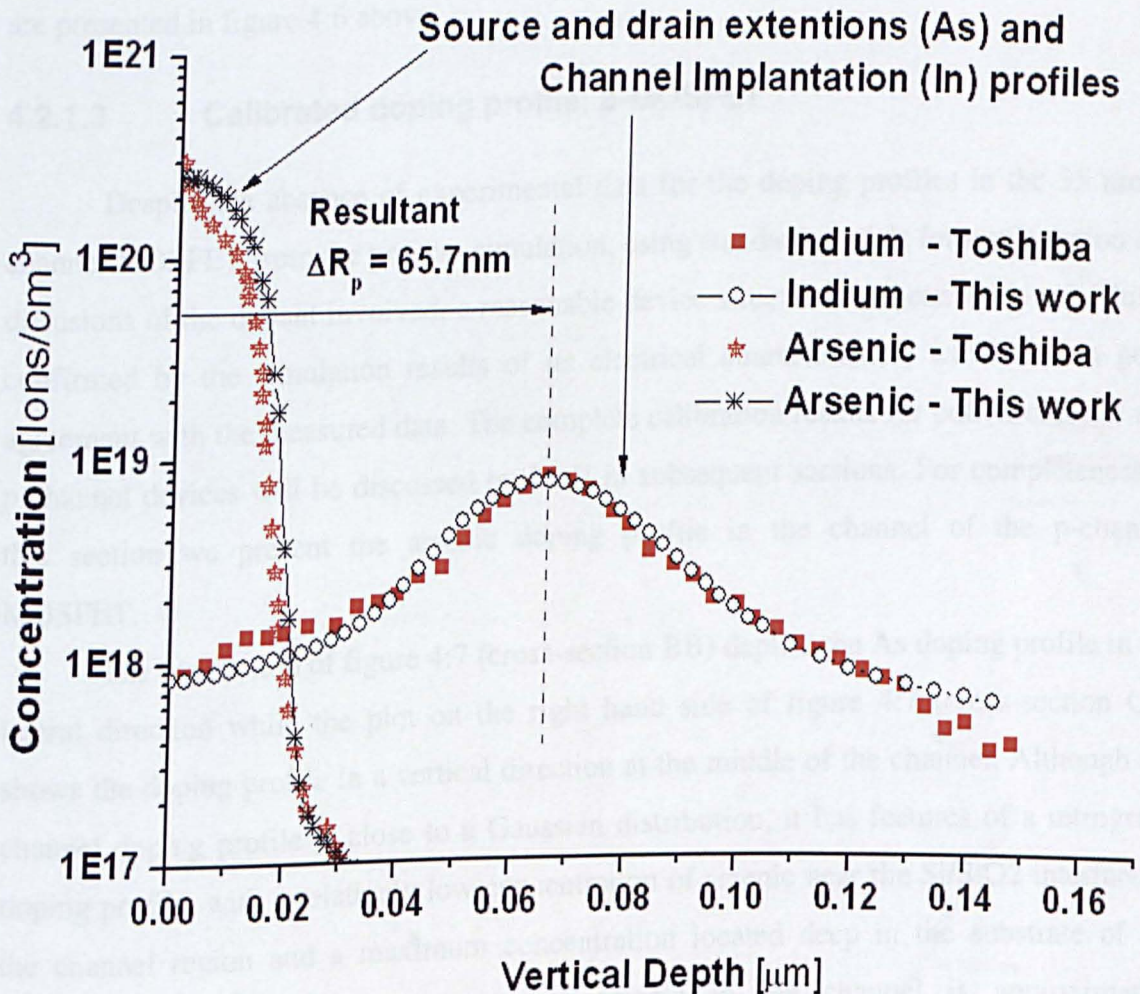


Figure 4:6 The calibrated doping profiles of the channel (indium dopants) and the source and drain extensions (arsenic dopants). The square and star symbols represent the measured data of indium and arsenic respectively. The corresponding simulation results are indicated by open circle and (*)

A continuous **In** doping profile in the channel region shown on figure 4:5 closely approximates the optimum from the design point of view, has a low-high (retrograde) step

doping profile illustrated with a dash line [4.5] [4.6] [4.7]. The higher **In** concentration in the substrate effectively suppresses short channel effects while the low concentration close to the Si/SiO₂ interface controls the threshold voltage and improves the mobility.

The shallow junction formation process involves a short thermal annealing time (~1 sec or less) and relatively high temperature (~1085°C) after the arsenic dopant implantation. The straightforward process simulation of the extensions gives a good agreement with the corresponding measured doping profile. The source and drain p-n junction at the extensions is about 20 nm deep. The fully calibrated process simulation results for the doping profile of indium in the channel and the As in the source and drain are presented in figure 4:6 above.

4.2.1.3 Calibrated doping profile: p-MOSFET

Despite the absence of experimental data for the doping profiles in the 35 nm *p*-channel MOSFET from the process simulation, using standard models for implantation and diffusions of the dopant involved, a reasonable device structure has been achieved. This is confirmed by the simulation results of its electrical characteristics, which are in good agreement with the measured data. The complete calibration results for both *n*-channel and *p*-channel devices will be discussed in detail in subsequent sections. For completeness in this section we present the arsenic doping profile in the channel of the *p*-channel MOSFET.

The top section of figure 4:7 (cross-section BB) depicts the As doping profile in the lateral direction while the plot on the right hand side of figure 4:7 (cross-section CC) shows the doping profile in a vertical direction at the middle of the channel. Although the channel doping profile is close to a Gaussian distribution, it has features of a retrograde doping profile, with a relatively low concentration of arsenic near the Si/SiO₂ interface in the channel region and a maximum concentration located deep in the substrate of the device. The maximum concentration of arsenic in the channel is approximately 2.3×10^{18} ions/cm³. As expected the extension junction depth (~28 nm) of *p*-channel device is bigger than that of the *n*-channel device (~20 nm).

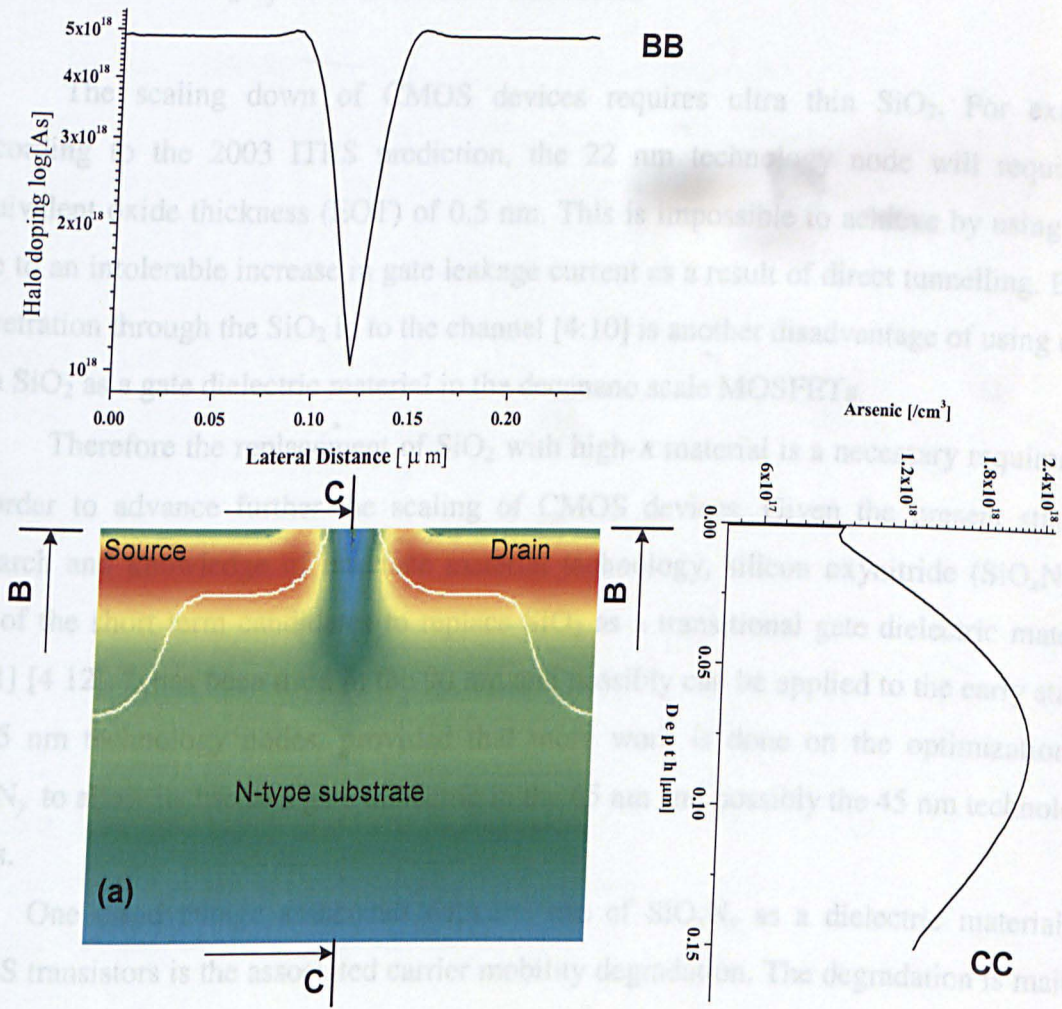


Figure 4:7 2-D illustration of continues doping of the channel profile of the *p*-channel MOSFET (a), 1-D doping profile of As in the lateral direction (cross-section BB) and the channel doping profile in the transverse direction (cross-section CC).

4.2.2 The halo doping and well implantation of the n-channel MOSFET

With the scaling of transistors to sub 50 nm gate lengths, the short channel effect (SCE) becomes a more dominant problem for conventional MOSFETs architectures. A negative threshold voltage shift with respect to the long channel devices, punch-through and an increase in *off*-current (sub- threshold current) are some of the adverse consequences of SCE in nano-scaled devices [4.8]. In order to reduce SCE, high dose retrograde channel doping and halo doping implantation techniques have been used in the process of channel engineering to manufactured the Toshiba 35 nm MOSFET [4.9].

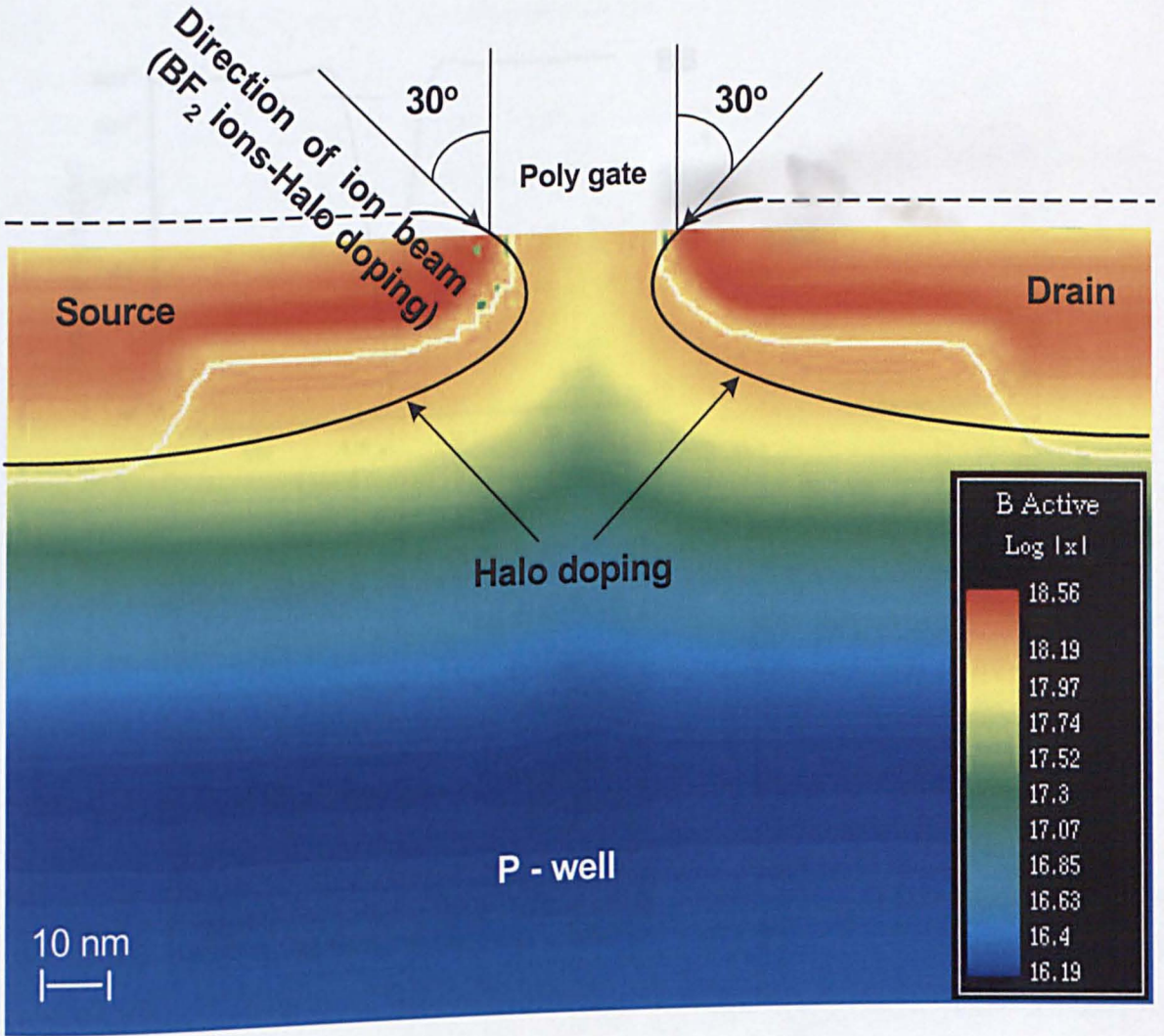


Figure 4:8 Halo doping profile of boron (B) the calibrated 35nm n-channel MOSFET.

BF_2 and As dopants have been used for halo implantation in the *n*-MOS and *p*-MOS devices respectively. As depicted on figure 4:8 (*n*-channel MOSFET) the ion-beam injection plane is tilted 30° off the vertical plane. The tilt angle gives a shape of the profile which circumscribes the source and drain extension regions bounded by the metallurgic *p*-*n* junctions. This will increase locally the doping concentration in the sub-gate region. As a result of this high concentration, punch-through will be suppressed by decreasing the source and drain depletion width.

4.3 SiO_xN_y as a dielectric material

The scaling down of CMOS devices requires ultra thin SiO_2 . For example according to the 2003 ITRS prediction, the 22 nm technology node will require an equivalent oxide thickness (EOT) of 0.5 nm. This is impossible to achieve by using SiO_2 due to an intolerable increase in gate leakage current as a result of direct tunnelling. Boron penetration through the SiO_2 in to the channel [4:10] is another disadvantage of using ultra-thin SiO_2 as a gate dielectric material in the decanano scale MOSFETs.

Therefore the replacement of SiO_2 with high- κ material is a necessary requirement in order to advance further the scaling of CMOS devices. Given the present state of research and knowledge of interface material technology, silicon oxynitride (SiO_xN_y) is one of the short term candidates to replace SiO_2 as a transitional gate dielectric material [4.11] [4 12]. It has been used in the 90 nm and possibly can be applied to the early stages of 65 nm technology nodes, provided that more work is done on the optimization of SiO_xN_y to allow its use as a gate dielectric in the 65 nm and possibly the 45 nm technology nodes.

One disadvantage associated with the use of SiO_xN_y as a dielectric material in CMOS transistors is the associated carrier mobility degradation. The degradation is mainly attributed to the penetration of nitrogen into the substrate region (oxynitridation) close to interface.

The process of the NO-oxynitride gate dielectric in the 35 nm Toshiba device has been optimized. This includes optimization through pre-growth surface engineering (cleaning) [4.13] and optimized fabrication techniques [4.14] [4.15], which minimize the penetration of nitrogen close to the interface. Indeed it has been shown that process-controlled oxynitridation can improve the electrical properties in comparison with the 'normal' SiO_2 [4.16] and enhance the high field mobility in n-MOSFET [4.17].

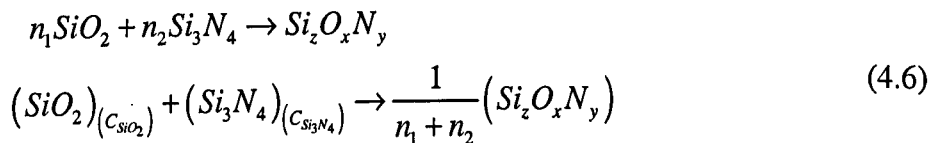
The dielectric constant of SiO_xN_y is not well documented. It varies generally with the mole fraction on the distribution of nitrogen in SiO_3N_4 . In the following section the estimation of the dielectric constant used in this work and the general relationship between the dielectric constant and the mole fractions of SiO_3N_4 and SiO_2 is presented.

4.3.1 Estimation of the SiO_xN_y dielectric constant

In this work SiO_xN_y was used as a gate dielectric material in the reference experimental device. Prior to performing the device simulation based on the device structure obtained from the process simulation, it is important to determine the dielectric constant of SiO_xN_y . As discussed in section 3.2.7, the permittivity of this material varies greatly with its stoichiometry. To estimate the dielectric constant of SiO_xN_y , the Clausius-Mossotti equation, which provides a quantitative relationship between the dielectric constant and polarisability of a material, has been adopted. The modified Clausius-Mossotti equation given in [4:18] is

$$\frac{\epsilon_r - 1}{\epsilon_r + 2} = C_A \frac{\epsilon_{rA} - 1}{\epsilon_{rA} + 2} + C_B \frac{\epsilon_{rB} - 1}{\epsilon_{rB} + 2} \quad (4.5)$$

where ϵ_r is the relative permittivity of a medium containing a mole fraction C_A of type A and mole fraction C_B of type B. To adapt equation 4.5 to the purpose of calculating the dielectric constant of silicon oxynitride, its chemical composition must be known. The chemical composition of SiO_xN_y is given by



From the principle of balancing the chemical equations and assuming that z is always one, the relationship between (n_1, n_2) and (x, y) shown in (4.7) can be established as

$$2n_1 = x, \quad 4n_2 = y, \Rightarrow n_1 + 3n_2 = z = 1 \quad (4.7)$$

Where n_1 and n_2 are mole numbers and (C_{SiO_2}) , $(C_{\text{Si}_3\text{N}_4})$ are mole fractions of SiO_2 and Si_3N_4 respectively given by the expressions

$$C_{SiO_2} = \frac{n_1}{n_1 + n_2} \quad ; \quad C_{Si_3N_4} = \frac{n_2}{n_1 + n_2} \quad (4.8)$$

During the calculation of the mole fraction of the individual compositions, it is assumed that the composition of SiO_xN_y is consistent with the oxidation states of silicon (+4), oxygen (-2) and nitrogen (-3).

Figure 4:9 shows the relationship between the dielectric constant (ϵ) of silicon-oxynitride and the mole fractions of silicon dioxide and silicon nitride. The values have been calculated using equations (4.5 - 4.8) and the stoichiometry of oxynitride. The dielectric constants of SiO_2 , $\epsilon_{ox} = 3.9$ and for pure nitride, $\epsilon_{Si_3N_4} = 7.9$ [4.19] have been used in the calculation.

The values (number of atoms) of the subscripts x and y in SiO_xN_y depicted in figure 4:9 are adopted from reference [4.20]. In this work, for device simulation of the 35 nm device model, the dielectric constant of silicon oxynitride have been calibrated to the value of $\epsilon = 5.45$, as indicated in figure 4:9. From figure 4:9 it can be estimated that the corresponding mole fraction of silicon dioxide and that of silicon nitride in the composition of the original experimental device are approximately 50%.

Making use of equations (4.5) and (4.7) one can determine the relationship between the mole fractions of the reactants (SiO_2 and Si_3N_4) and the subscripts x and y in the product (SiO_xN_y).

$$\begin{aligned} C_{SiO_2} &= \frac{n_1}{n_1 + n_2} = \frac{3x}{2(x+1)} \\ C_{Si_3N_4} &= \frac{n_2}{n_1 + n_2} = \frac{y}{2(2-y)} \end{aligned} \quad (4.9)$$

Therefore, it is possible from equations 4.8 and 4.9 to approximate the amount of oxygen (x) and nitrogen (y) that have been used in the fabrication of the actual experimental device. At the same time one can perform a linear interpolation using the information from the graph in figure 4:9 and the dielectric constant, $\epsilon = 5.45$, obtained

from the calibration process. By performing a simple linear interpolation on the curve shown in figure 4:9, the composition of oxygen and nitrogen has been calculated to be 0.48 and 1.01 respectively.

Using the equations in (4.9) and mole fractions of silicon dioxide and silicon oxynitride, the respective calculated values are 0.51 and 0.994. It is, therefore, possible to conclude that the likely overall composition of oxygen and nitrogen, which have been used to manufacture the experimental device of the 35 nm transistor is $\text{SiO}_{0.48}\text{N}_{1.01}$.

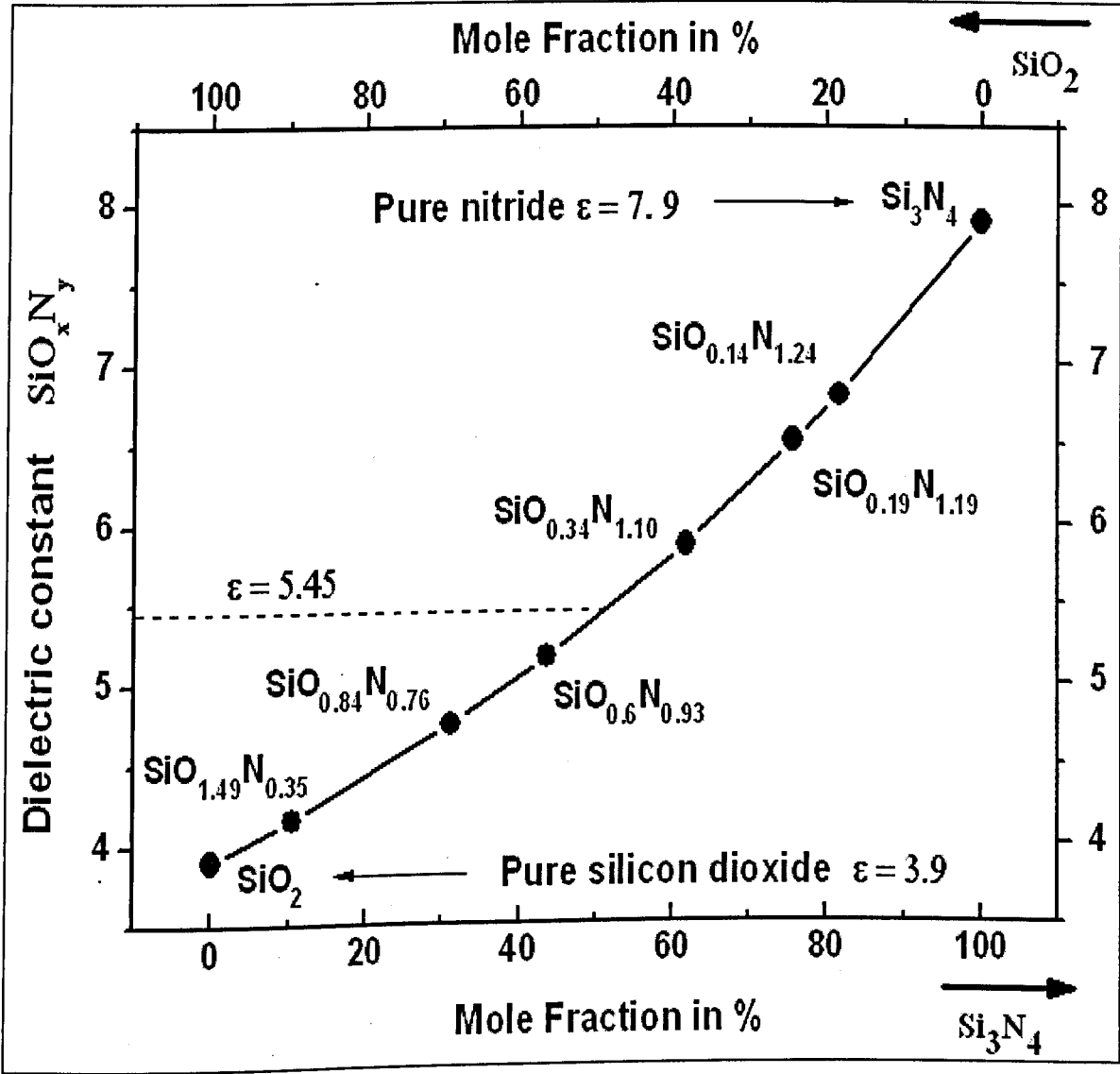


Figure 4:9 Relation of dielectric constant of silicon oxynitride to the mole fraction of silicon dioxide and silicon nitride calculated using the Clausius-Mossotti equation.

4.4 Calibration of electrical parameters at low and high V_{dd}

In the previous sections the estimation of the channel, source and drain doping profiles of the 35 nm prototype transistor have been discussed. The doping profiles are very important in the overall calibration process, determining the electrostatic behaviour, carrier transport in the channel, and the access resistances needed in electrical simulations. The calibration of the process simulation sequence was a difficult task because of the interaction of many process parameters. Once the device structure has been established the next stage in the calibration is the matching of important electrical parameters; such as threshold voltage, subthreshold slope, on and off current, and the overall current-voltage characteristics.

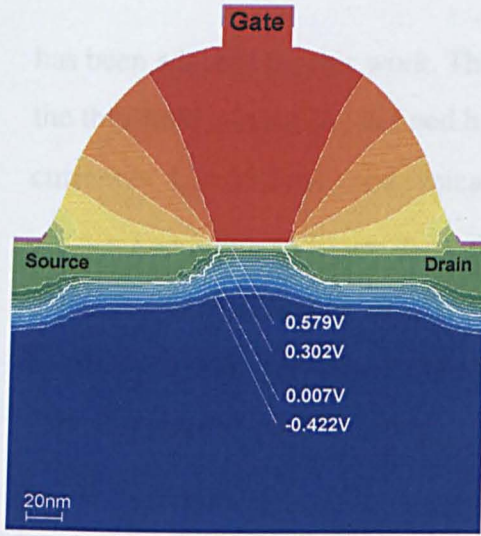
This section describes the calibration of the device simulations which has been performed in order to match the simulated current-voltage characteristics of the prototype 35 nm MOSFET with the experimental data. The results of the calibration of the n and p -channel MOSFETs are presented in the next two subsections.

The systematic calibration methodology, discussed briefly in section 4.1 has been applied to device simulation of the 35 nm transistor. This calibration enables us to optimize critical electrical model parameters used in the simulation of both n -channel and p -channel MOSFETs by using the structure obtained from the full process simulation. The simulation results are then verified against the real device properties.

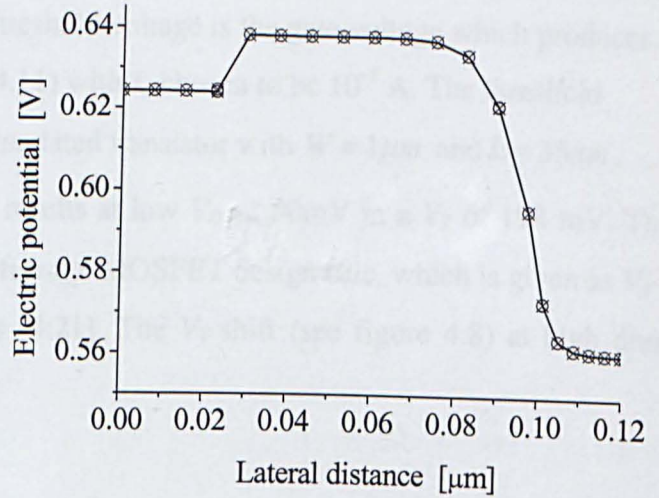
The first step in the calibration process is to match the subthreshold characteristics of the calibrated device structure with the measured data. This step determines the electrostatic integrity of the device structure and any corresponding short channel effects. In the subthreshold region the current is exponentially dependent on the height of the potential barrier between the source and the drain (see figure 4:10 b), which is controlled by the gate voltage and influenced by the short channel effects.

In the subthreshold regime, the electrons (in the case of n -channel MOSFET) are injected from the source over this barrier more into the drain end of the channel. The carrier concentration is low and the Poisson equation is decoupled from the current continuity equation. The source end of the barrier height is entirely determined by the solution of the Poisson equation and is very sensitive to the doping distribution in the

channel. The careful matching of the subthreshold characteristics verifies the 2D doping distribution in the transistor and is a stepping stone for the successful calibration at high current conditions.



(a)



(b)

Figure 4:10 2D electric potential in the simulated 35nm n-channel MOSFET at $V_G=0V$ and 1D potential profile across the channel

One of the most important MOSFET parameters that has to be matched in the calibration is the threshold voltage (V_T). The simplest definition of V_T is the gate voltage which results in strong inversion producing the same concentration of electrons at the interface as that of the whole concentration in the bulk and is given by:

$$V_T = V_{fb} + 2\psi_B + \frac{\sqrt{2\epsilon_{si}qN_a2\psi_B}}{C_{ox}} \quad 4.10$$

where V_{fb} is the flat band voltage ($V_{fb} = \Phi_{ms} - \frac{Q_f}{C_{ox}}$), Q_f , a fixed charge at the interface and ψ_B is the Fermi potential. The threshold voltage defined by (4.10) cannot be measured directly and is not suitable for practical applications. It also does not take into account the short channel effect.

A threshold voltage criteria based on subthreshold current dependant empirical formula given by

$$I_T(V_g = V_T) = I_o \frac{W}{L}, \quad 10^{-8} \leq I_o \leq 10^{-6} \quad 4.11$$

has been adopted for this work. The threshold voltage is the gate voltage which produces the threshold current (I_T) defined by (4.11) with I_o chosen to be 10^{-7} A. The threshold current is $I_T = 28.5\mu\text{A}$ for a typical simulated transistor with $W = 1\mu\text{m}$ and $L = 35\text{nm}$.

According to figure 4:10, this results at low V_D of 50mV in a V_T of 198 mV. This value of V_T is reasonable, since it satisfies the MOSFET design rule, which is given as $V_T \leq 0.25V_{dd}$ for high performance devices [4:21]. The V_T shift (see figure 4.8) at high drain voltage is approximately 70mV.

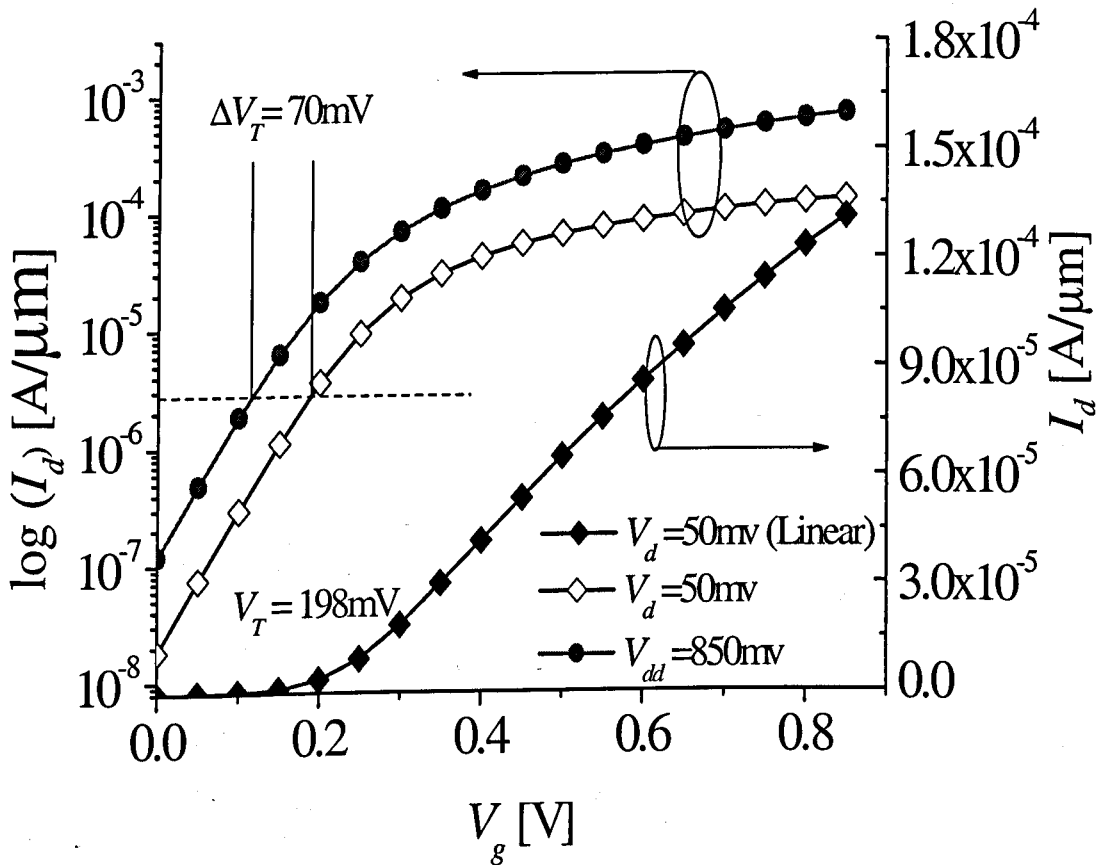


Figure 4:11 Threshold voltage estimation using the current criteria.

The next step in the calibration process is to match the low drain voltage characteristics of the transistor above threshold, these being affected by the low field mobility and the series resistance. This matching involves adjustment of the channel mobility and vertical field dependence of the mobility. The contact resistances are also introduced at this stage. Care should be taken when adjusting the values of the contact resistances, which have similar effect as the vertical field dependence of the mobility on the I_d - V_g characteristics at low V_d .

The final step in the calibration process is matching of the device characteristics at high drain voltage in the saturation regime. As previously mentioned, the subthreshold characteristic is dominated by the electrostatic integrity of the device. Good control of the channel doping is required to minimize short channel effects. The I_d - V_g characteristics at low V_d are dominated by the low field mobility and the series resistance. The saturation region is dominated by non-equilibrium transport including velocity overshoot, modelled in the DD approach by the lateral field dependence of the mobility and also by the electric field behaviour in the neighbourhood of pinch-off. The high electric field has a significant effect on channel velocity, as reflected in the mobility model. This will directly affect one of the main device performance indicators, the drive current (I_{on}).

In the real 35 nm device, a relatively high I_{on} with supply voltage of 850mV has been achieved through process optimization. This optimization, which includes process induced strain, may not be captured sufficiently by the simulation models, which have been discussed in chapter 3. Given the importance of the high electric field effect and strain on carrier transport, it is necessary during the calibration process to adjust default mobility parameters (see section 4.5) in the concentration dependent mobility model in order to *match* the simulated result for I_{on} to the experimental value.

Generally, the most difficult part of the calibration process is to achieve the optimum device structure from the full process simulation, which can satisfy the requirements for both subthreshold and saturation regimes during the device simulation. As illustrated in the flow chart showing the calibration methodology (figure 4:1), it is important to loop between the full process and device simulation procedures until a satisfactory device structure has been achieved. This was a delicate balancing act between

meeting the requirements for high performance and low subthreshold leakage current without compromising the model parameters at both extremes of device operation .

4.4.1 Simulation results for the well calibrated 35 nm n-MOSFET

In this sub section, the results of the calibration of the n-channel 35 nm gate length MOSFET are presented. Figures 4.12 and 4.13 depict the I_d - V_g characteristics in both logarithmic and linear scale. Figure 4.14 shows the overall output characteristics (I_d - V_d) of the calibrated 35nm n-MOSFET simulations compared to the experimental data of the Toshiba device.

There is no experimental data available for low I_d - V_g at low drain voltage. The data shown in figures 4:12, 4:13, 4:14 and 4:16 was extracted from the published I_D - V_D characteristics at $V_D = 50\text{mV}$. It was difficult in particular to extract values for $V_G < V_T$. However, by taking all the results into account, the simulated electrical characteristics of the prototype device are in good agreement with the experimental measurements.

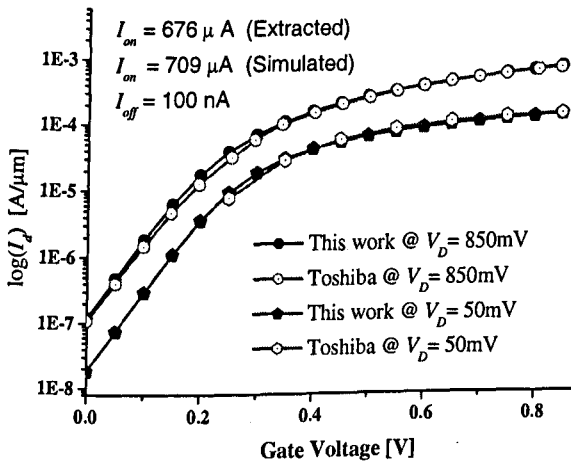


Figure 4:12 Well calibrated I_d - V_g characteristics of 35 nm n-channel MOSFET at a high drain voltage of 850mV

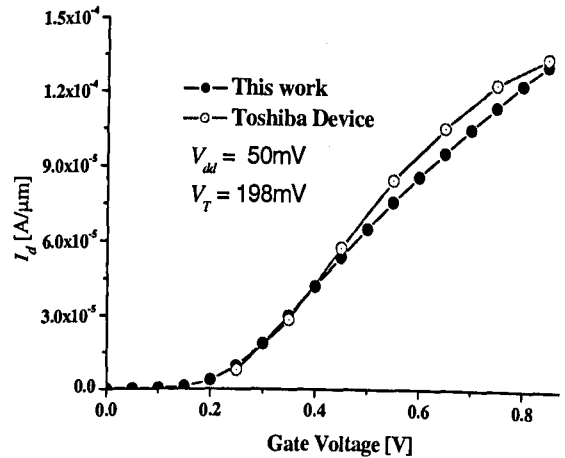


Figure 4:13 Well calibrated I_d - V_g characteristics of the 35 nm n-channel MOSFET at a low drain voltage of 50mV

The simulations deliver a saturation drive current of $I_{on} = 709 \mu\text{A}/\mu\text{m}$, which is within a 5% error margin from the measured data. The subthreshold leakage current in both the simulated and the experimental cases is close to $I_{off} = 100 \text{ nA}/\mu\text{m}$.

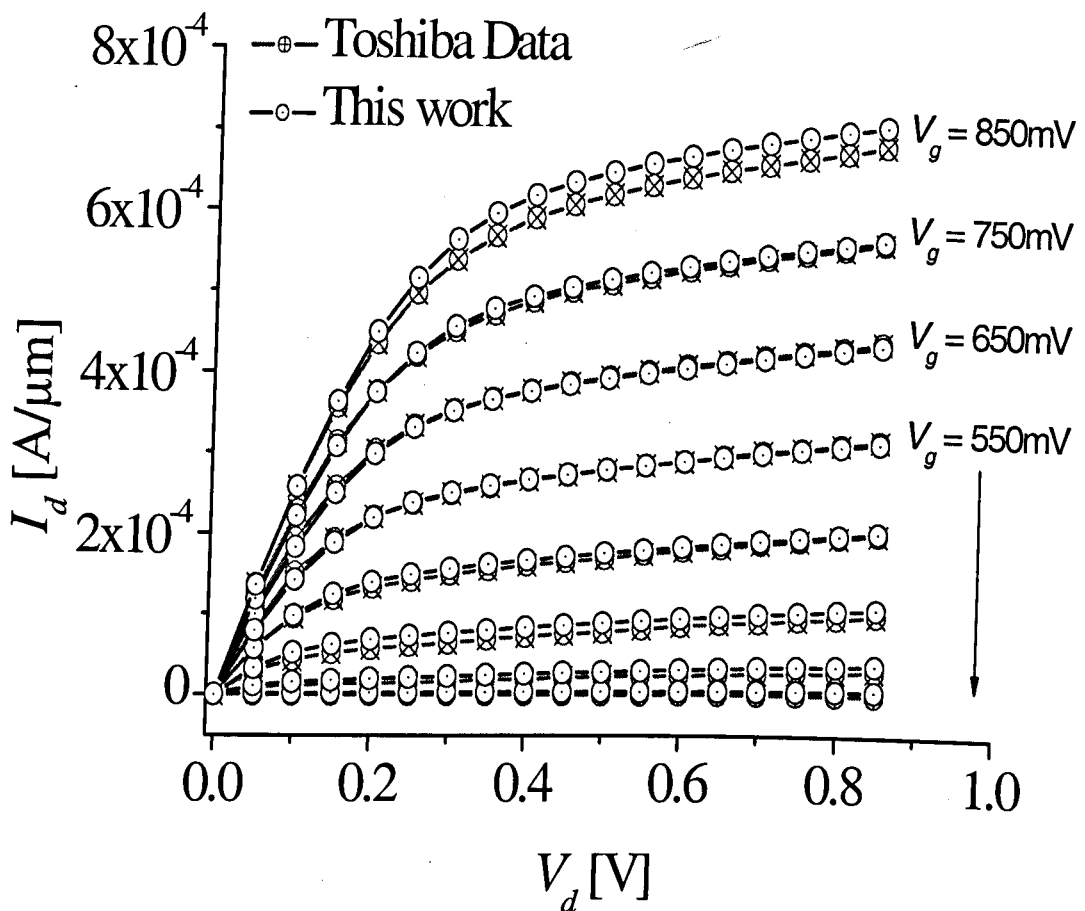


Figure 4:14 Output characteristics of the calibrated 35nm n-channel MOSFET (I_d - V_d)

4.4.2 Simulation results for the calibrated 35nm p-channel MOSFET

Although there is no available measurement data for the channel doping profile of the p-channel MOSFET, good calibration has been achieved based on the published electrical characteristics of the experimental device and personal communication with the authors. Moreover, the process simulation in the p-channel case was facilitated by the fact that only dopants (P and As) with well defined properties in Taurus were used in the fabrication process. In this sub section the calibration and simulation results are presented.

Figures 4:15 and 4:16 depicts the I_d - V_g characteristics of the p-channel calibrated model transistor which is tuned with the experimental data at the high and low drain voltages of $V_{dd} = 850\text{mV}$ and $V_d = 100\text{mV}$ in both logarithmic and linear scale. Figure 4.17 shows the overall output characteristics (I_d - V_d) of the same device for different gate voltages.

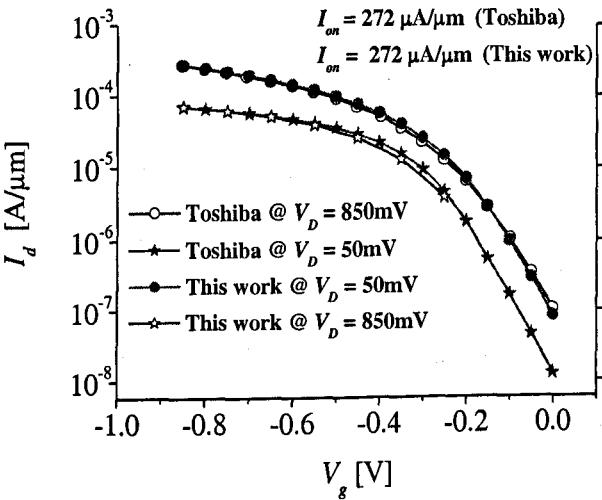


Figure 4:15 I_d - V_g characteristics of the calibrated 35nm p-channel MOSFET on a logarithmic scale.

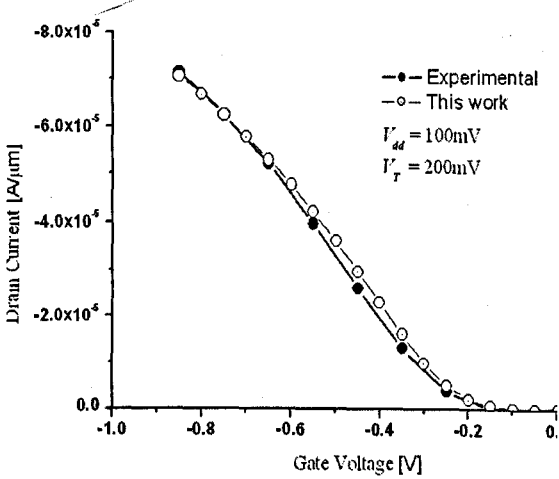


Figure 4:16 I_d - V_g characteristics of the calibrated 35 nm p-channel MOSFET at low V_d

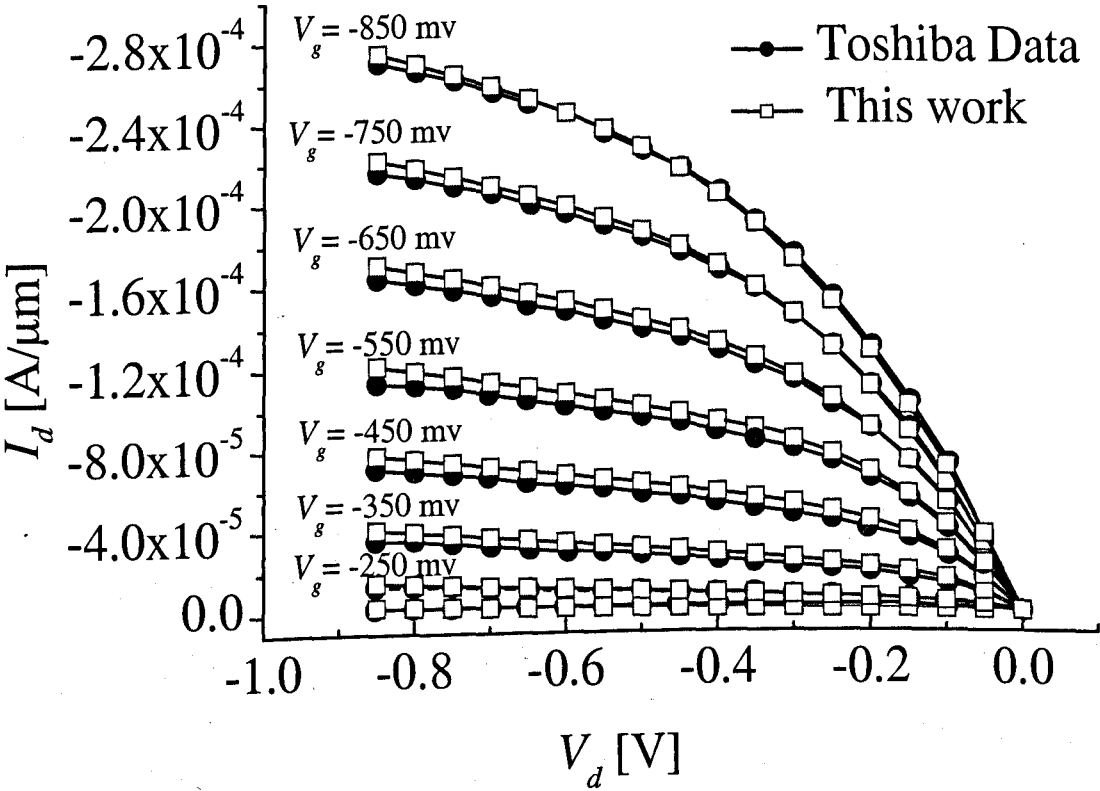


Figure 4:17 Output characteristics (I_d - V_d) of the calibrated 35 nm p-channel MOSFET. Each curve has been simulated for a corresponding fixed gate voltage (V_g) and variable drain voltage (V_d).

The agreement between the calibrated simulation results and the experimental data for the *p*-channel MOSFETs is very good in both the I_d - V_g (see figure 4:15 and 4:16) and the overall output I_d - V_d characteristics (see figure 4:17). A very good agreement of the simulated drive current, $I_d = 273 \mu\text{A}/\mu\text{m}$, has been achieved compared to the experimental value of $I_d = 272 \mu\text{A}/\mu\text{m}$.

4.5 Mobility and electric field in the calibrated 35 nm n-MOSFET

Carrier mobility is one of the important parameters which directly influences the device performance. Even in the modern decanano MOSFETs where non-equilibrium transport effects play an important role, there is a 50% correlation between the channel mobility and the drive current [4.22]. The performance of the conventional transistors which are required for 90 nm technology node and beyond is highly degraded due to the poor carrier mobility in the channel. The carrier mobility degradation can be attributed mainly to the high channel doping and associated ionised impurity scattering [4.23] [424]. The corresponding vertical field also reduces the surface mobility as a result of increased surface roughness scattering [4.25] according to the universal mobility curve [4.26] (see figure 4:18).

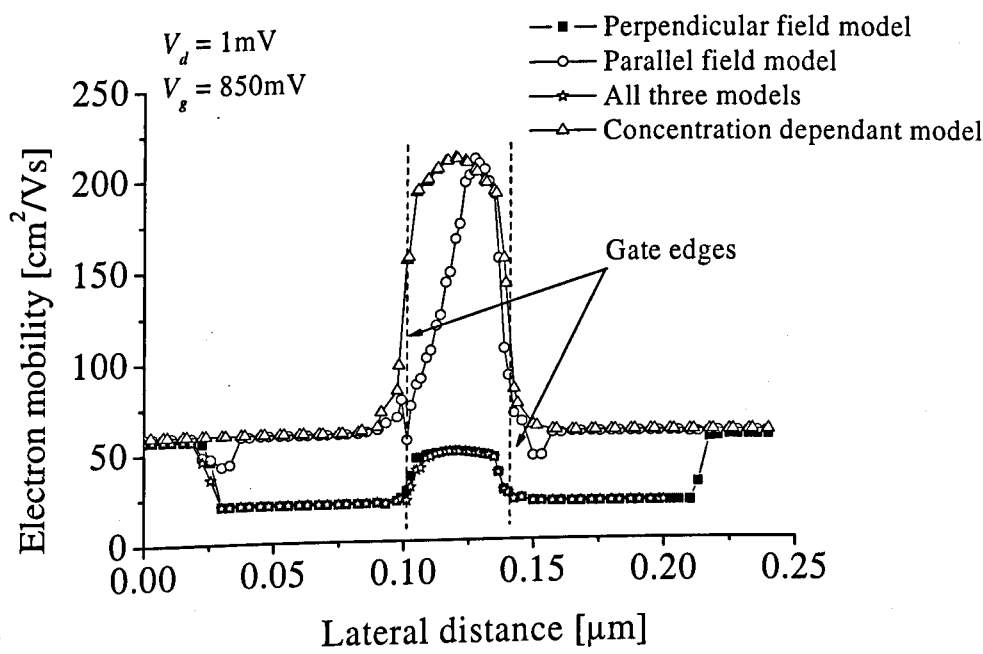


Figure 4:18 The effect of different mobility models on the carrier mobility in the 35 nm prototype transistor.

Here we examine closely the electric field and carrier mobility in the channel of the calibrated 35 nm gate length n-channel MOSFET at low and high drain voltages of 50mV and 850mV respectively. A comparison between the average channel mobility obtained using the default and modified parameters of the previously determined mobility models in Taurus device is presented.

Figures 4:19 and 4:20 show the channel carrier mobility in the simulated 35 nm n-MOSFET at low drain voltage of 50mV and at the supply voltage of 850mV respectively. The solid line represents the mobility calibrated to the Toshiba device by adjusting the parameters of the carrier mobility model and the thin line represents the default Si-parameters in Taurus.

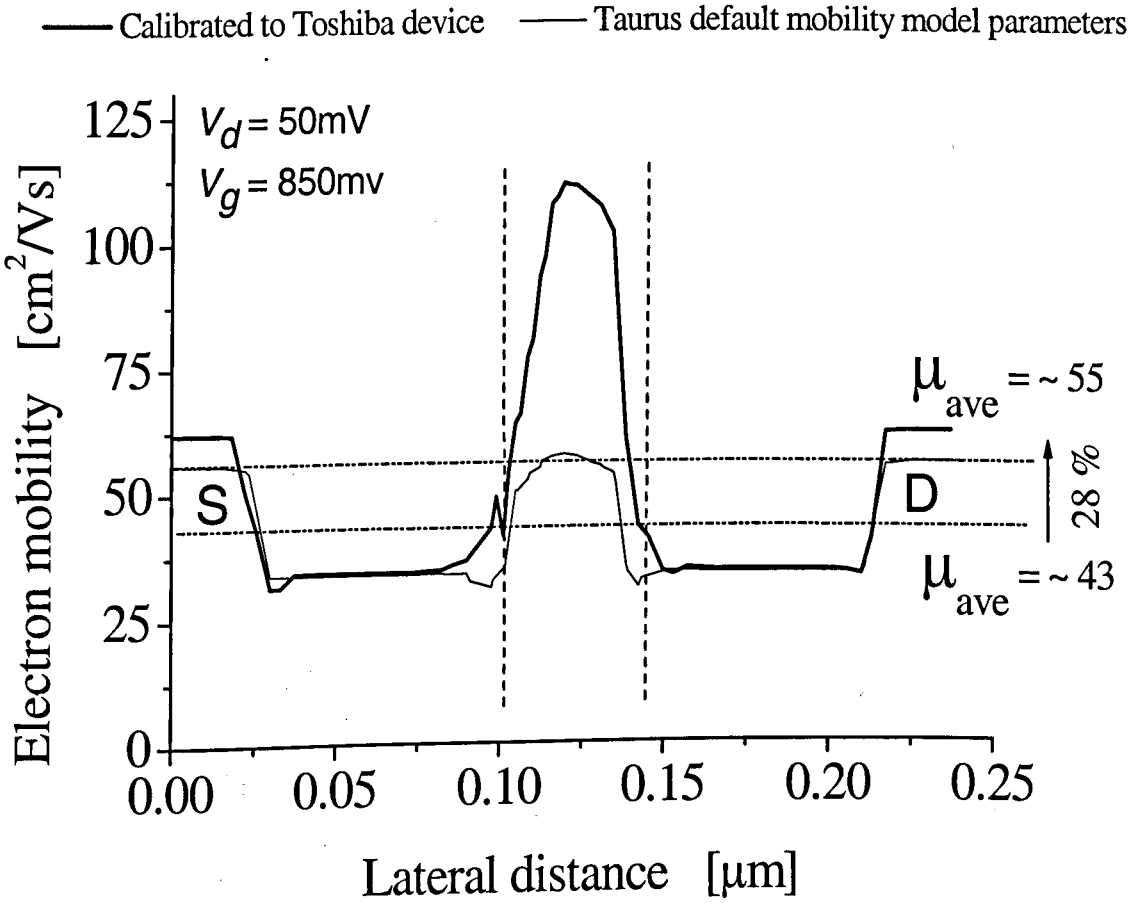


Figure 4:19 Carrier (electron) mobility versus the lateral distance across the channel at a low drain voltage of 50mV. The dashed lines in the vertical direction indicate where the gate edges are located.

At a low drain voltage of 50mV, the average default carrier mobility in the channel is $\sim 43 \text{ cm}^2/\text{V}\cdot\text{s}$. When the modified mobility model parameters are introduced, the channel carrier mobility increases to $55 \text{ cm}^2/\text{V}\cdot\text{s}$, which is an increase of $\sim 28\%$ (see figure 4:18). The corresponding carrier mobility in the channel degrades further when a high drain voltage of 850mV is applied. In these conditions there is 57% difference between the average default and modified mobility in the channel (see figure 4:20)

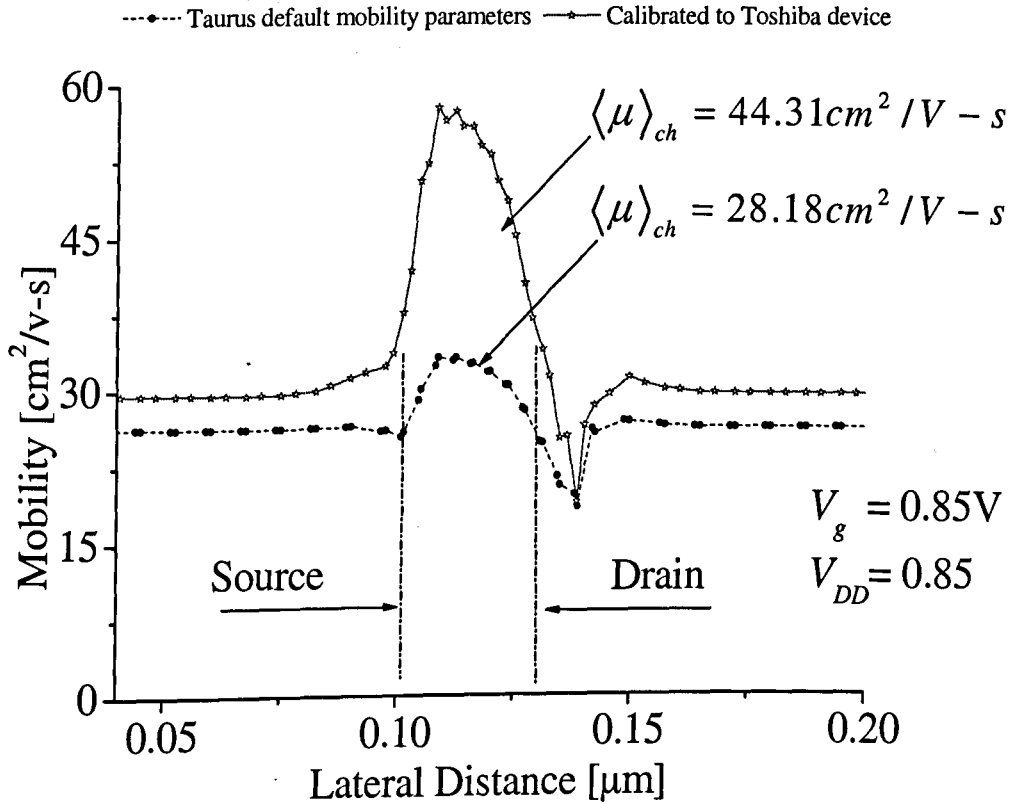


Figure 4:20 Carrier (electron) mobility versus the lateral distance across the channel at the high drain voltage of 850mV. The dashed lines in the vertical direction indicate where the gate edges are located.

This means, that in order to match the performance of the experimental n-channel device, which delivers $676 \mu\text{A}/\mu\text{m}$ drive current, the mobility model parameters have to be tuned to achieve an approximately 57% mobility enhancement. This is due to the inadequacy of the default mobility model in describing the universal mobility curve in the presence of high channel doping as the DD approach fails to capture non-equilibrium transport phenomena in the simulated devices.

Moreover, during the fabrication of the 35 nm transistor, process induced strain contributes to the mobility and device performance enhancement. The SiN_4 contact etch stop layer introduces a tensile stress on the top gate layers, which propagate down to the channel area and increase the channel mobility [4.27]. A Monte Carlo device simulation study on n-channel Toshiba MOSFET also shows that the device may have a process-induced strain equivalent to 5% Ge content which is necessary to reproduce the experimental data [4.28]. Therefore it is essential, from a simulation point of view, to tune the mobility parameters to the tolerable degree of error discussed in chapter 3.

4.6 Device structure and parameters of the calibrated 35nm MOSFET

The structure and device parameters obtained as a result of the calibration of the 35 nm gate length Toshiba MOSFET are presented in this section. Figure 4:21 shows the net 2D doping profile and the principal design dimensions of n-channel MOSFETs together with the typical domain of simulation.

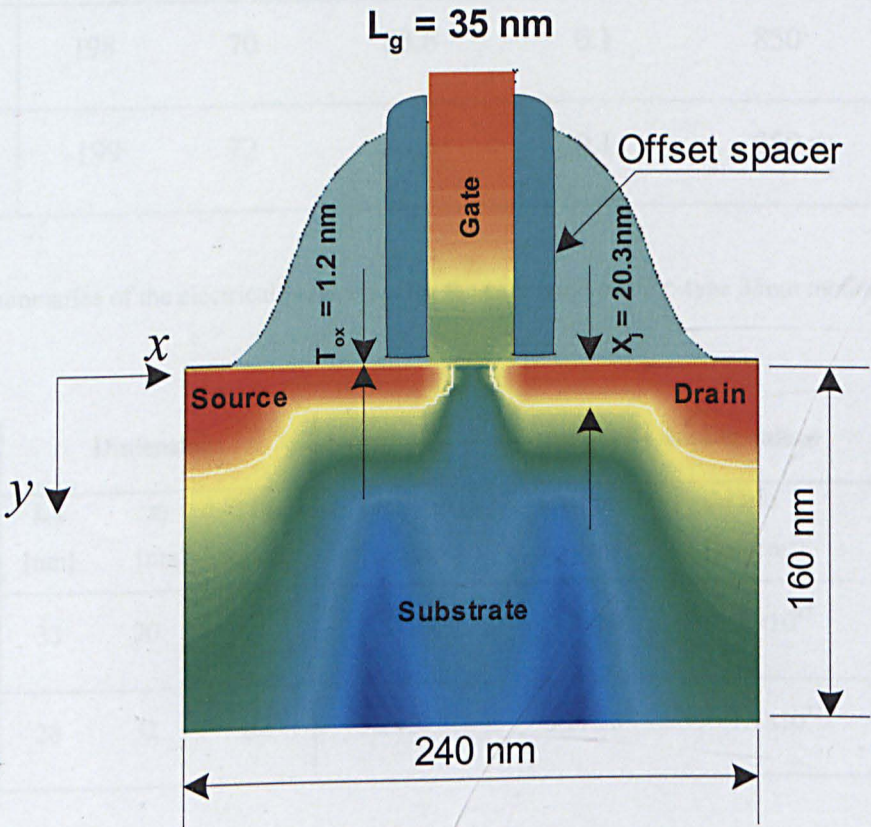


Figure 4:21 2D net doping profile and important design parameters of the calibrated 35 nm gate length model.

The dimensions of the p-channel MOSFET are similar to that of the n-channel transistor with the exception of the vertical junction depth (j_x), which is wider in the p-channel MOSFET due to fast boron diffusion. The junction depth of the real p-MOSFET is in the region of 30-35 nm and the calibrated value is 32 nm. The typical 2D simulation domain is 240 nm \times 160 nm which *is* enough to accommodate the depletion region around the drain at the supply voltage. The adopted axis orientation is also indicated in figure 4:21. For three dimensional simulations the device width is defined in the ‘z’ direction.

The important design and device dimensions and electrical parameters of both the n and p channel devices are given in tables 4:4, and 4:5. Thus values have been used as initial (reference) parameters for the subsequent scaling process which is described in the next section. ΔV_T in table 4.4 is the threshold voltage difference measured at high and low drain voltages.

	V_T [mV]	ΔV_T [mV]	S [mV/dec]	I_{off} [μ A/ μ m]	V_{dd} [mV]	I_{on} [μ A/ μ m]
<i>n</i> -MOS	198	70	83.6	0.1	850	709
<i>p</i> -MOS	-199	72	84.5	0.1	850	273

Table 4:4 Summaries of the electrical parameters for the calibrated n and p-type 35nm model MOSFET

	Dimensions			Impurity doping values			
	L_{ch} [nm]	x_j [nm]	t_{ox} [nm]	Extension [ion/cm ²]	Channel [ion/cm ²]	Halo [ion/cm ²]	Back ground [ion/cm ³]
<i>n</i> -MOS	33	20	1.2	8×10^{15}	3.5×10^{13}	9×10^{12}	1×10^{17}
<i>p</i> -MOS	28	32	1.2	2×10^{15}	2.5×10^{13}	1.7×10^{13}	1×10^{17}

Table 4:5 Summaries of the important device dimensions and impurity doping values of the calibrated n and p-channel of the calibrated 35 nm y MOSFETs

4.7 Scaling Strategy

The overall scaling strategy is based on the generalised scaling rules which are reviewed in chapter 2 [4:29]. In addition to the generalised scaling rule the technology roadmap (ITRS'03 edition) has been adopted as a guide for important device parameters including I_{off} and I_{on} requirements up to the year 2018. Table 4:6 and 4:7 show some of the selected near and long term technology node characteristics (for more roadmap data which are relevant to this work, see chapter 2). This section is devoted to results of the scaling of the 35 nm transistor according to the ITRS requirements.

The first part details the scaling scenarios which have been followed. The second part discusses in detail the steps followed during the device and process simulations. In the final part of this section the electrical properties of the scaled devices (including I - V characteristics at high and low drain voltage, off and on current threshold voltage and the subthreshold slope) are presented. This is complimented by a comparative analysis of the net channel doping, the carrier mobility and the electric field in the scaled and the prototype devices.

4.7.1 The scaling scenarios of the conventional MOSFET devices

The well calibrated 35 nm gate length n-channel MOSFET which has been discussed in the previous sections of this chapter, has been used as a reference device for the scaling of transistors required for the next four technology node (65, 45, 32, and 22 nm) generations. The physical gate lengths of the corresponding transistors are 25, 18, 13, and 9 nm respectively. These devices represent the dimensions and the requirements for high performance transistors up to the end the ITRS (see table 4:6 and 4:7). The physical gate length has been taken as minimum features size in the process of scaling.

Only the conventional MOSFET architecture is considered in the scaling process. The needs for technology boosters in terms of strain and mobility enhancement are highlighted in the scaling process to compensate for the well known deficiencies of conventional MOSFETs.

Year of Production	2003	2004	2005	2006	2007	2008	2009
Technology Node		hp 90			hp 65		
DRAM ½ Pitch [nm]	100	90	70	70	65	57	50
MPU Printed L _G [nm]	65	53	45	40	35	32	28
MPU Physical L _G [nm]	45	37	32	28	25	22	20

Table 4:6 Near term predictions of technology road map for semiconductors: (ITRS 2003 edition)

Year of Production	2010	2012	2013	2015	2016	2018
Technology Node	hp 45		hp 32		hp 22	
DRAM ½ Pitch [nm]	45	35	32	25	22	18
MPU Printed L _G [nm]	25	20	18	14	13	10
MPU Physical L _G [nm]	18	14	13	10	9	7

Table 4:7 Long term predictions of technology road map for semiconductors (ITRS 2003 edition)

4.7.2 Doping in the scaled devices

With the scaling of conventional MOSFETs over the next generations of technology nodes, the design and the manufacture of the optimum channel, source and drain doping profiles becomes increasingly difficult. Specifically, the formation of ultra shallow source and drain extensions (for the 35 nm gate length CMOS and beyond) needs massive effort in order to obtain the optimum junction depth and resistivity. The junction depths of the devices, which are the subject of this work are in the range of 5 – 20 nm. Since the error tolerances are very small, the parameters describing the doping distributions should be selected very carefully.

Shallow junction formation involves implantation and post implantation rapid thermal annealing or other advanced activation steps with low thermal budget, in order to activate the required doping. Aside from shallow junction formation, the ion implantation into the channel and the extension regions also require the same level of attention.

To assist the design of the doping distributions in the scaled devices we investigate the scaling properties of the doping distribution in the reference 35nm. The effect of the two scaling factors κ and α on the ion projection range, standard deviation, and dose of implantation are studied in the next section. They are used only as guidance to the scaling process.

4.7.2.1 Scaling properties of R_p and σ

In chapter 2, we discussed the general scaling theory and the impact of the corresponding two scaling constants, κ and α on device dimensions, doping concentration and key electrical parameters. How do the two scaling factors affect the parameters? In this section we deduce the relationship between the two scaling factors (κ and α) and the parameters such as the projection range, the standard deviation and the dose of implantation that describe the doping distribution in the scaled devices.

We expect the projection range to shrink in the same manner as the vertical and horizontal physical dimensions of the device linearly by a factor of $1/\kappa$. Therefore, we can establish the relationship between the newly scaled value of R_p' and the original R_p as

$$R_p' = \frac{1}{\kappa} R_p \quad (4.12)$$

Making use of equation (4.12) we can write the relation of the standard deviation of the ion distribution in the scaled device to the standard deviation of the ion distribution in the originally calibrated prototype 35nm device as

$$n\left(\frac{x}{\kappa}\right) = n_o \exp \left[-\frac{\left(\frac{x}{\kappa} - \frac{R_p}{\kappa}\right)^2}{2\sigma_s^2} \right] \quad (4.13)$$

Here, in expression in (4.13), we assume that the doping profile shifts toward the Si/SiO₂ interface proportionally to $1/\kappa$ and the distribution narrows κ times. We can simplify expression (4.13) to obtain the expression for standard deviation of the doping profile in the scaled device (σ_s) by rearranging the terms and assuming that the maximum doping concentration in the scaled device increases by $\alpha\kappa$ compared to the original device giving $n_0 = \alpha\kappa n_0$, and $n\left(\frac{x}{\kappa}\right) = \alpha\kappa n(x)$. Therefore, the new standard deviation of the scaled device is then becomes

$$\sigma_s = \frac{1}{\kappa} (R_p - x) \left(2 \ln \left| \frac{n(x)}{n_0} \right| \right)^{-1} \quad (4.14)$$

Comparing the expression of the new and the original standard deviations we arrive in the intuitive expression $\sigma_s = \sigma/\kappa$. Therefore both the projection range and the lateral struggling will be shrunk in the scaling process linearly in the same way as the device dimensions. The resulting channel doping profiles in the scaled devices are illustrated in figure 4:22.

This approach to the scaling of the doping profile preserves the important element of the doping design of the original 35 nm MOSFET. The maximum concentration of indium at the interface in each of the scaled devices remains between $1.0 - 5.0 \times 10^{18}/\text{cm}^3$. The maximum doping concentration is in the range of $0.8 - 3 \times 10^{19}/\text{cm}^3$, and is shifted towards the interface as the device gate length becomes smaller. The increases in the doping concentration in the channel region (particularly the interface) create an unfavourable environment for carrier transport. The corresponding electron mobility degradation produces low drive current and is a major concern in the scaling of conventional MOSFETs.

In addition to the concerns related to carrier mobility degradation in the scaled devices due to high doping concentration, there is another important issue associated with *solid solubility*[†] limit of indium in Si. Its solid solubility limit is in the range of $1 - 1.5 \times 10^{19}$

[†] Solid-solubility shows the degree of which one solid (crystal) component can dissolve another material to form a coherent solution.

atoms/cm³ provided that laser annealing is used for doping activation [4.30]. The required maximum concentration of indium in the 9 and 13nm scaled transistors (see figure 4.15) is above this limit. This will be an additional scaling limitation factor from the processing point-of-view, unless new advanced non-equilibrium doping techniques are developed.

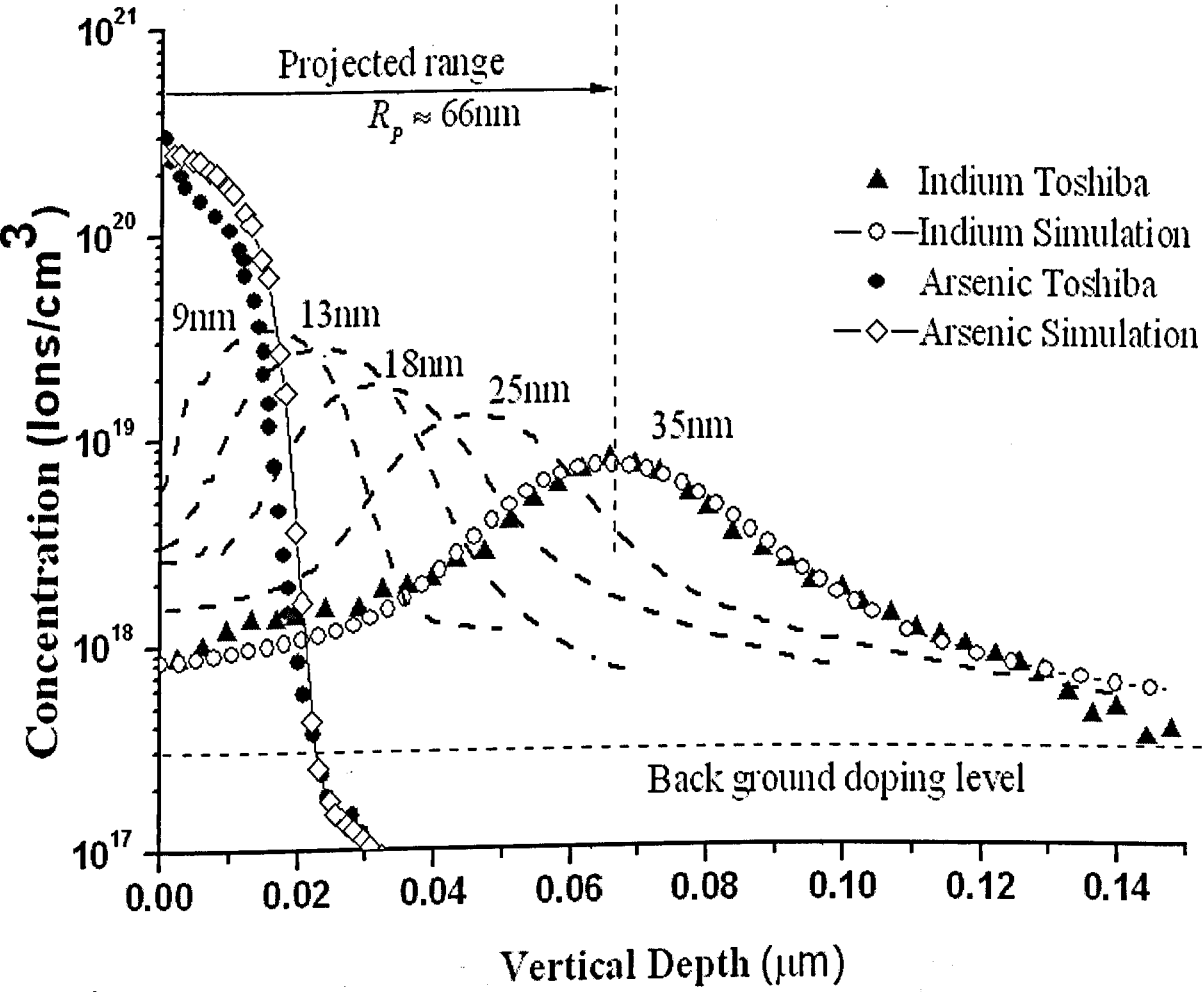


Figure 4:22 Doping profiles of the channel region in the reference 35 nm and the scaled devices of 25, 18, 13, and 9nm gate length. The dashed lines show the channel doping profile of the scaled devices. The solid triangle and star symbols represent the experimental data for indium channel doping and arsenic doping in extensions respectively. The background doping ($\sim 3 \times 10^{17}$) is marked by the horizontal dashed line.

4.7.3 Scaling results: Electrical properties of the scaled devices

Among the most important MOSFET electrical parameters are the threshold voltage, the off-current and the nominal drive current. Therefore, in each case of the scaled devices the *on* and *off* currents have been carefully compared to the ITRS requirements. The focus is on high performance transistors specified in the ITRS. There is, however, a difference in the supply voltage specified by the current edition of the ITRS and the supplied voltages used in this work as shown in table 4:7 which corresponds to the 2001 edition of the ITRS.

The main reason for this difference is the fact that the original 35 nm Toshiba MOSFET used for the calibration has a supply voltage of $V_{dd} = 850$ mV. The supply voltage of the scaled devices has been obtained by scaling of this reference supply voltage. Although the increase of the supply voltage to match the ITRS[‡] values increases I_{off} and I_{on} simultaneously, these differences do not affect the overall scaling trend investigated in this work.

Although with careful design of the doping profile the off-current requirement of the ITRS could be achieved with conventional MOSFET scaling, the achievement of the on-current requirements is a very difficult task due to the channel doping, even if strain enhancement of the mobility is assumed.

Table 4:9 shows the average mobility values, which are obtained from the device simulation of the reference 35 nm, and that of scaled 25, 18, 13, and 9 nm gate length devices. The values given in column three and four are the average channel mobilities using default and modified mobility parameters respectively. The adjusted parameters are adopted to represent mobility enhancement due to the introduction of a reasonable amount of strain in the channel. The table gives a clear indication of what is the percentage of mobility enhancement in the conventional MOSFETs in order to meet the ITRS requirements. We did not evaluate the possibility of introduction of new channel materials like Ge or III-V semiconductors.

[‡] The ITRS data shown in table 4:7 are from the latest 2003 edition which considers the introduction of SOI and other mobility enhanced devices.

L_g [nm]	Supply Voltage (V_{dd}) [V]		Drive current (I_{on}) [μ A/ μ m]		Off-current (I_{off}) [nA/ μ m]	
	This work	ITRS	This work	ITRS	This work	ITRS
35/37	0.85	1.2	670	980	100	50
25	0.8	1.1	1200	1510	140	70
18	0.7	1	1356	1900	210	100
13	0.6	0.9	1153	2110	260	300
9	0.5	0.8	900	2400	550	500

Table 4:8 Comparative values of supply voltage, drive current and off current for the reference 35nm and scaled 25, 18, 13 and 9nm devices

L_g [nm]	V_{dd} [mV]	Average channel mobility		$\Delta\mu$ [%]
		With default parameters [$cm^2/V\text{-s}$]	with modified parameters [$cm^2/V\text{-s}$]	
35	850	28	44	57
25	800	21	36	71
18	700	19	35	84
13	600	18	32	78
9	500	13	21	62

Table 4:9 Average channel mobility values in the reference and scaled devices at the corresponding supply voltages. $\Delta\mu$ represents the percentage difference between the simulated carriers mobility using modified parameters (column 4) and the default parameters (column 3) provided in the Taurus device simulation mobility parameters.

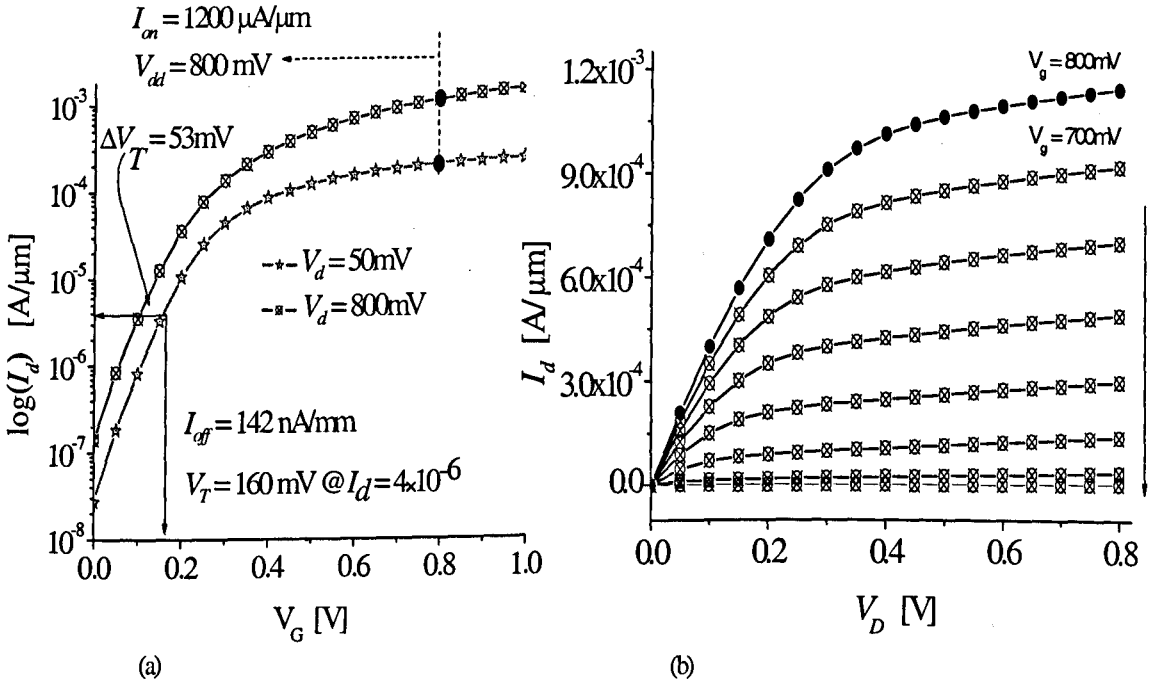


Figure 4:23 I_d - V_g characteristics of the scaled 25 nm MOSFET (a) and its overall output characteristics (I_d - V_d). The dotted curve on I_d - V_d characteristics corresponds to the gate voltage which is equivalent to the supply voltage of 800mV.

The I_d - V_g and the I_d - V_g characteristics of the 25nm MOSFET which corresponds to the 65nm technology node are illustrated in figures 4.22a and 4.22b respectively. The on current at drain voltage of $V_d = 0.8\text{V}$ is $1200 \mu\text{A}/\mu\text{m}$. The drive current required for the 65 nm technology node according to ITRS prediction is $1500 \mu\text{A}/\mu\text{m}$ at a drain voltage of $V_{dd} = 1.1\text{V}$.

Extrapolating the values of the drain current in the simulated I_d - V_g characteristics to gate and supply voltages of 1.1 V, it can be seen that the ITRS requirement for the on current can be achieved in this device. However, the I_{off} requirement will be compromised by increasing the supply voltage, introducing power constraint problems.

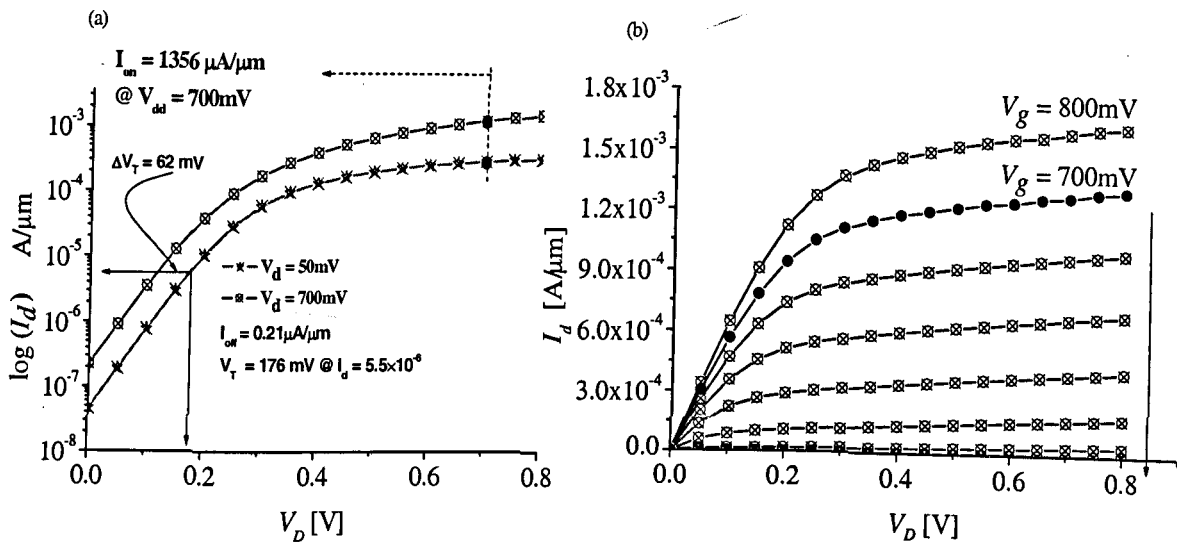


Figure 4:24 I_d - V_g characteristics of the scaled 18 nm MOSFET (a) and its overall output characteristics (I_d - V_d). The dotted curve on I_d - V_d characteristics corresponds to the gate voltage which is equivalent to supply voltage of 700 mV.

Similarly figures 4:24a and 4:24b depict the I_d - V_g characteristics and the I_d - V_d characteristics for the 18 nm transistor respectively, which corresponds to the 45 nm technology node generation. While the drive current is $1356 \mu\text{A}/\mu\text{m}$ with a supply voltage of 700 mV, the ITRS requirement of $1900 \mu\text{A}/\mu\text{m}$, may be achieved at the prescribed supply voltage of 1 V. The simulated value of the on-current under such condition is $2228 \mu\text{A}/\mu\text{m}$, which is about 24% more than the ITRS requirement (table 4:7).

Figures 4:25a and 4:25b show the I_d - V_g and the I_d - V_d characteristics of the scaled 13 nm transistor respectively. The solid symbols in figure 4.25b correspond to a supply voltage of 600 mV. The corresponding drive current is $I_{on} = 1153 \mu\text{A}/\mu\text{m}$. The ITRS requirement for the drive current in this device is $I_{on} = 2110 \mu\text{A}/\mu\text{m}$ at a supply voltage of $V_{dd} = 900 \text{ mV}$. The simulation shows that the 13 nm device could deliver a drain current of $I_d = 2276 \mu\text{A}/\mu\text{m}$ at gate voltage of $V_g = 900 \text{ mV}$, which is 8% more than the ITRS requirements. Increasing the supply voltage, however, will increase the subthreshold leakage current.

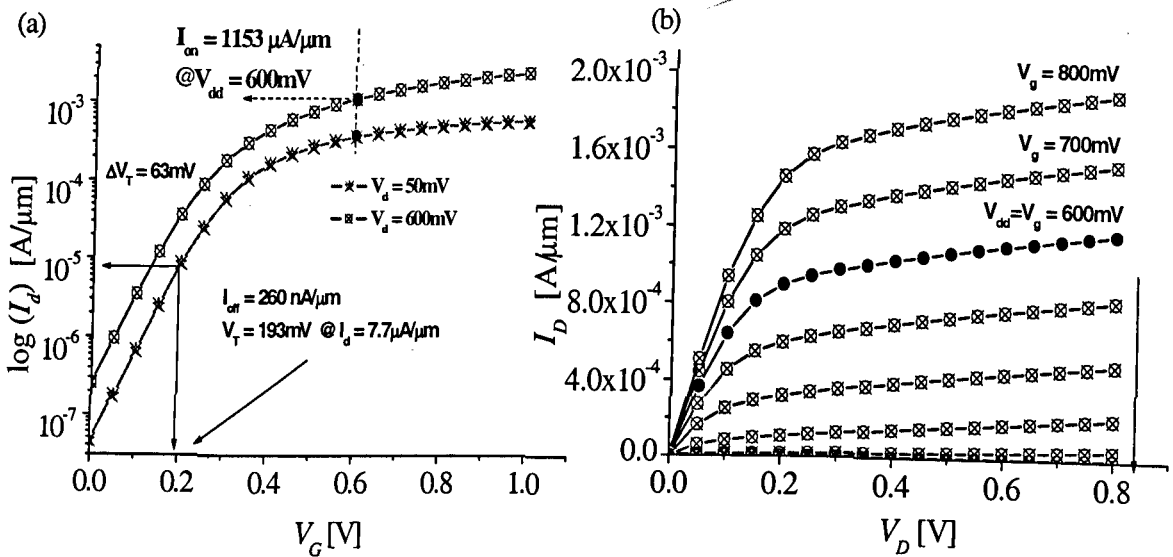


Figure 4:25 I_d - V_g characteristics of the scaled 13 nm MOSFET (a) and its overall output characteristics (I_d - V_d). The dotted curve on I_d - V_d characteristics corresponds to a gate voltage, which is equivalent to a supply voltage of 600mV.

Continued scaling of conventional MOSFETs at its present pace will become increasingly difficult as we approach the end of the ITRS time line. The 22 nm technology node requires a 9nm gate length MOSFET which will deliver a corresponding drive current of $I_d = 2400 \mu\text{A}/\mu\text{m}$ and $I_{off} = 500 \text{ nA}/\mu\text{m}$. However, the simulation of this device shows that this target is very difficult to achieve.

Figures 4:26a and 4:26b illustrate the I_d - V_g and I_d - V_d characteristics respectively. Assuming the present doping arrangements used in this project, the simulations of the 9 nm MOSFET gives drives a current of $1000 \mu\text{A}/\mu\text{m}$, which is 58% less than the ITRS requirements for such devices.

The simulated value of the *off*-current is fairly close to the ITRS data (see table 4:8). This I_{off} match in the 9nm MOSFET has been achieved by applying a high level of channel doping (the indium concentration at the interface is $\sim 6 \times 10^{18}/\text{cm}^3$, figure 4:22). The fact that the interface is highly doped means significant degradation of device performance manifested as low drive current.

Increasing the supply voltage from 0.5V (used in this work) to 0.8V (ITRS specification) may significantly increase I_{on} to 2086 $\mu\text{A}/\mu\text{m}$. Despite the increase in drive current, however, the sensitive subthreshold current is significantly increased, which leads to a higher power dissipation.

Moreover the level of threshold voltage, $V_T = 229\text{mV}$, compared with the supply voltage of $V_{dd} = 500\text{mV}$ is unacceptable. For an optimum device operation, $V_T \leq (0.2 - 0.25)V_{dd}$ is considered to be acceptable from a device design point of view.

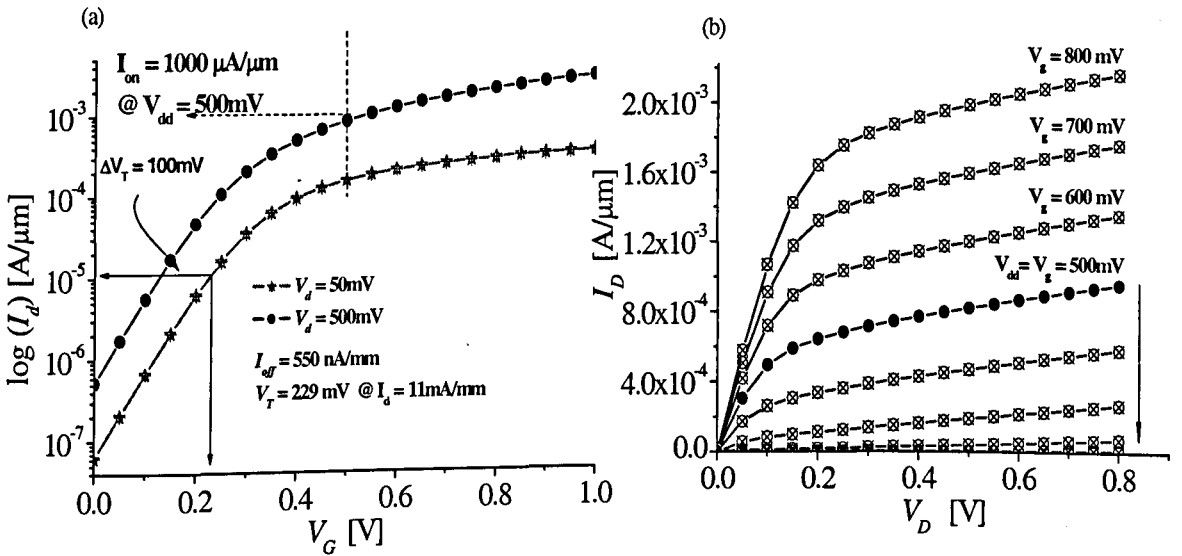


Figure 4:26 I_d - V_g characteristics of the scaled 9 nm MOSFET (a) and its overall output characteristics (I_d - V_d) The dotted curve on I_d - V_d characteristics corresponds to a gate voltage which is equivalent to a supply voltage of 500mV.

In addition to the high doping concentration in the channel that causes performance degradation in deeply scaled decanano MOSFETs, the possible inadequacy of the DD approach must also be flagged. It is known that the DD model may not be a suitable simulation tool to capture accurately the non-equilibrium carrier transport in such small scale devices.

4.8 Comparative study of the calibrated and scaled device properties

In this section we present comparisons of the properties and the electrical characteristics of the five transistors, *i.e.*, the well calibrated 35 nm Toshiba MOSFET and the four scaled device. Figure 4:27 depicts the I_d - V_g characteristics of all five devices. The data are plotted both in logarithmic (left axis) and linear scale (right axis), so that the sub threshold and the saturated regions are well illustrated. Excellent electrostatic integrity is achieved in the scaled devices with the subthreshold practically the same as in the original 35 nm prototype transistor.

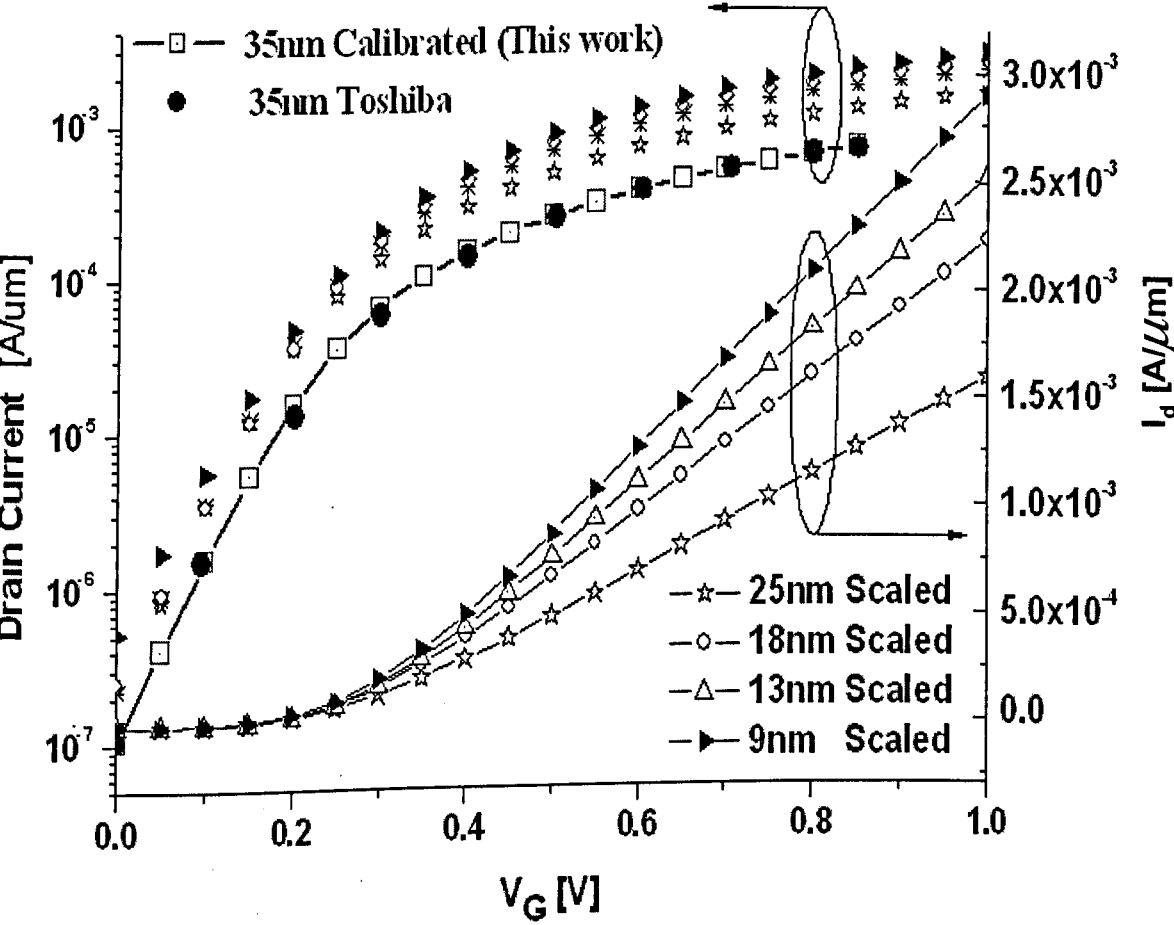


Figure 4:27 Drain current as a function of gate voltage on both a logarithmic and linear scale to highlight the electrostatic properties and drive current in the saturated regime respectively. The logarithmic scale includes the experimental data for the reference 35 nm Toshiba MOSFET and calibrated device.

Some of the drive currents and off-currents reported in the literature, together with the ITRS requirements and simulation results from this work are illustrated in figure 4:28a (with the corresponding supply voltages laid out in figure 4:28b). Given the present process technologies and device architecture, both the I_{on} and I_{off} requirements of the ITRS latest edition seem too difficult to be adequately met for the 65 nm technology node and beyond.

What the industry requires for future generation technology nodes, continually increasing drive current, increasingly difficult, as the gate length approaches towards the 22 nm technology node both from the results from our work and reported data in the literature (see figure 4:28a). In fact the ITRS anticipates the introduction of new device architectures and materials to deliver the requirements outlined in its 2003 edition. Notice, however, that the simulation results of the drive current and the off-current from the scaled devices reasonably reflect the reported set of data as shown in figure 4:28a.

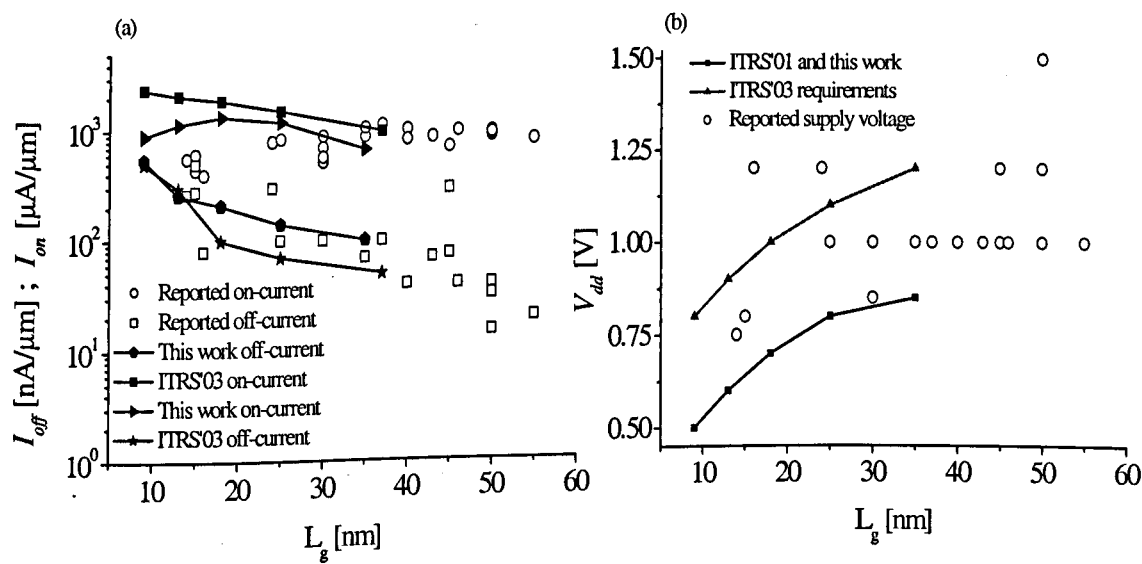


Figure 4:28 Comparison of the gate length with the off and on-currents of MOSFETs from reported data, ITRS'03 requirement trends and simulation results of this work (a) and the corresponding drain current (b). Reference [4.31-4.39]

The supply voltages depicted in figure 4:28b used in the simulation of the scaled devices are similar to the specification given in ITRS 2001 edition. The original Toshiba device which was used to calibrate the prototype 35 nm transistor was fabricated to meet the ITRS 2001 edition requirements. Hence the supply voltage values for the scaled MOSFETs are inherited from this edition.

Figure 4.28 shows the electric field in the prototype and the scaled devices. The electric field in the channel increases in agreement with the scaling prediction $E' \rightarrow \alpha E$, with decreasing gate length [4:21]. The field increases because the supply voltage scales much slower than the device dimensions. At such a high field velocity overshoot is expected to contribute the increased performance in the scaled devices compensating some of the negative effects of the high doping concentration in the channel. The DD simulation used for this project does not capture the velocity overshoot effect, and therefore underestimates the device performance.

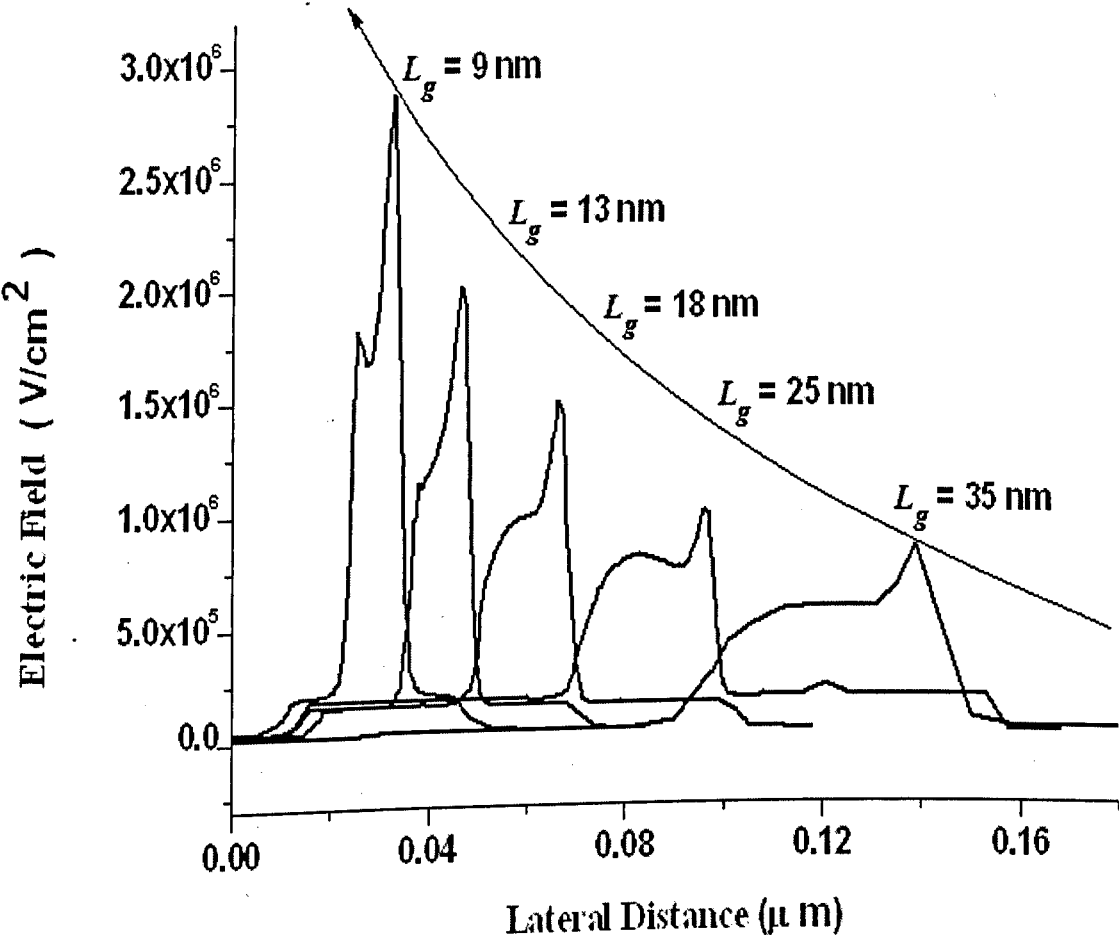


Figure 4:29 The electric field profile across the channel just below the Si/SiO₂ interface. The peak points in each device occurs near the drain edge of the p-n junction.

Figure 4:30 shows the net doping concentration along the channel in the prototype 35nm and the scaled transistors. The net doping concentration in the channel reaches as high as $2 \times 10^{19} / \text{cm}^3$ in the case of the 9 nm gate length device, which is well above the solid solubility limit of silicon. High doping is essential in order to minimise the short channel effects and to control punch through (using halo doping) in decanano MOSFETs. While this is needed in order to keep the electrostatic integrity of devices, it creates unfavourable conditions for carrier mobility which leads to performance degradation.

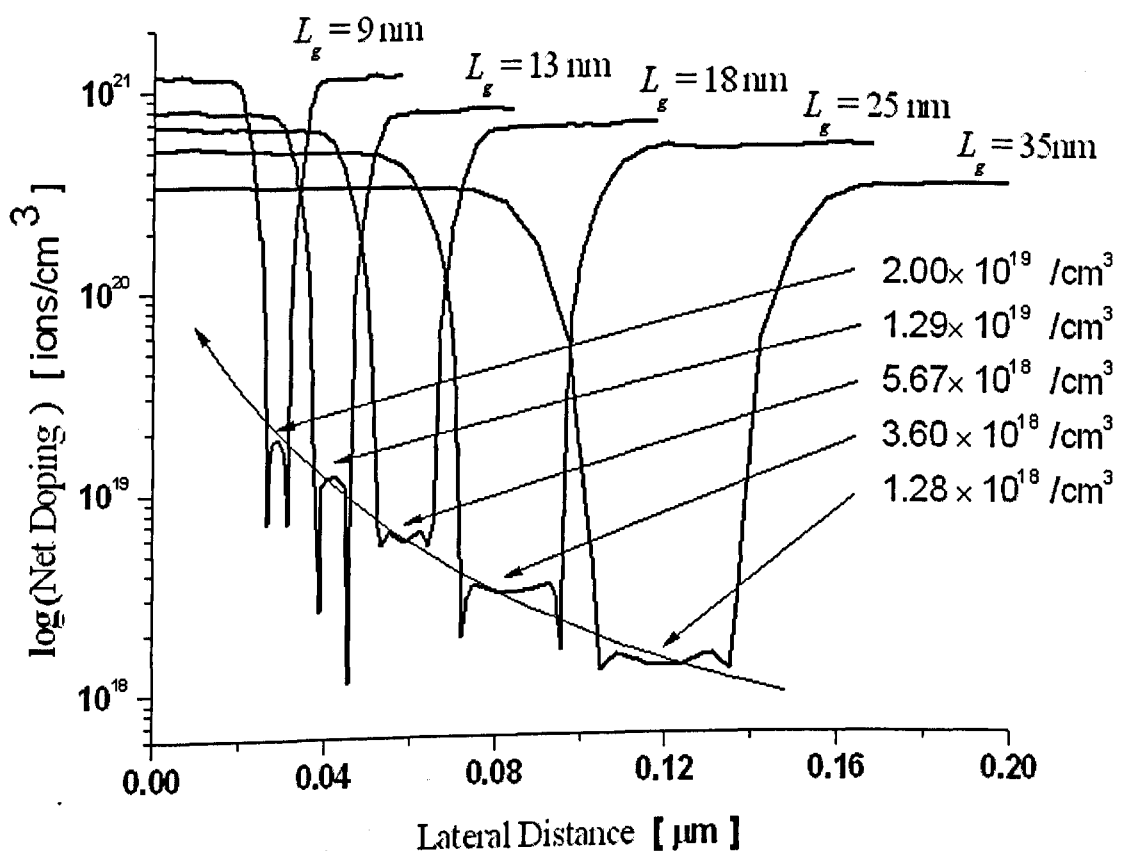


Figure 4:30 Net doping concentration in the channels of the reference 35 nm and the scaled 25, 18, 13, and 9nm gate length devices. The concentration points indicated by the arrows are located in the geometric centre of the corresponding devices.

4.9 Conclusion

In this chapter, the scaling of conventional MOSFETs that corresponds to the next generations of technology nodes have been presented. The scaling is based on systematic calibration of the simulation tools and methodology with respect to the 35 nm *n* and *p*-channel prototype MOSFETs. In order to achieve a realistic device structure full scale process simulation was carried out for the scaled transistors and the reference 35 nm MOSFETs. The output of the process simulation was used in the electrical simulations of the corresponding devices.

The calibration of both the *n*-channel and *p*-channel prototype MOSFETs was successful. The experimentally measured channel doping of the *n*-channel MOSFET was accurately reproduced by process simulation. The simulated current-voltage characteristics of the *n* and *p*-MOSFETs are in very good agreement with the experimental measurements. The agreement between the measured and the simulated device characteristics also validates the device structures obtained from the process simulation.

The subthreshold leakage current $I_{off} = 100\text{nA}/\mu\text{m}$ and the reasonably high drive current $I_{on} = 709\mu\text{A}/\mu\text{m}$ for a supply voltage of $V_{dd} = 850\text{mV}$ were well reproduced. The simulations also show that the transport in the prototype device has been enhanced as a result of strain induced by a balanced deposition of Si_3N_4 used as a contact etch stopping layer.

Based on the prototype 35 nm MOSFET, scaling of conventional MOSFETs, which are required by the road map up to the year 2018, have been performed using simulation. Excellent electrostatic integrity has been achieved for the scaled devices.

At the same time the high channel doping needed to suppress short channel effects results in significant carrier mobility and device performance degradation. This is a real problem when it comes to the current drive requirements for all the devices required by the ITRS. I_{on} cannot be achieved under standard Si mobility parameters especially considering that mobility models predict a need for carrier mobility enhancement by up to 78 % in the case of 13 nm transistor.

Another important issue which must be taken in to consideration is the solid-solubility limit of dopants in crystalline silicon. In particular, near the end of the latest ITRS time line, the doping concentration in conventional MOSFETs is well above the typical solid solubility limits for most dopants in silicon.

4.10 Chapter references

-
- [4:1] P. Bodrogi, and T. Tarczali, "Investigation of colour memory", in *Colour Image Science*, Edited by Lindsay W. MacDonald, and Ronnier Luo, *John Wiley & sons, LTD*, pp. 56, 2002
- [4:2] T. Johnson, "Methods of characterizing colour scanners and digital cameras", in *Colour Engineering*, Edited by Phil Green, and Lindsay MacDonald, *John Wiley & Sons, LTD*, pp. 245, 2003
- [4:3] S. Inaba, K. Okano, S. Matsuda, M. Fujiwara, A. Hokazono, K. Adachi, K. Ohuchi, H. Suto, H. Fukui, T. Shimizu, S. Mori, H. Oguma, A. Murakoshi, T. Itani, T. Inuma, T. Kudo, H. Shibata, S. Taniguchi, T. Matsushita, S. Magoshi, Y. Watanabe, M. Takayanagi, A. Azuma, H. Oyamatsu, K. Suguro, Y. Katsumata, Y. Toyoshima, and H. Ishiuchi, "High Performance 35 nm Gate Length CMOS With NO Oxynitride Gate Dielectric and Ni Salicide", *IEDM Tech. Dig.*, pp. 641 - 644, December 2001
- [4:4] S. M. Sze, *Semiconductor Devices, Physics and Technology*, 2nd Edition, John Wiley & Sons, INC, New York, 2002
- [4:5] M. E. Law, "Process modelling for future technologies", *IBM J. RES. & DEV*, Vol. 46 (2,3), pp. 2002
- [4:6] J. B. Jacobs and D. Antoniadis, "Channel Profile Engineering for MOSFET's with 100nm Channel Lengths", *IEEE Trans. Electron Devices.*, Vol. 42, No. 5, pp. 870, May 1995
- [4:7] I. De and C. M. Osburn, "Impact of Super-Step-Retrograde Channel Doping Profile on the Performance of Scaled Devices", *IEEE Trans. Electron Dev.*, Vol. 46, No. 8, pp. 1711-1717, 1999
- [4:8] R. F. Pierret, *Field Effect Devices, Volume IV, Modular series on Solid State Devices*, 2nd Edition, Addison-Wesley Publishing Company, 1990
- [4:9] Y. Taur, C. H. Wann, and D. J. Frank, "25nm CMOS Design Consideration ", *IEDM Technical Digest*, pp. 789, 1998
- [4:10] M. L. Green, E. P. Gusev, R. Degraeve, and E. L. Garfunkel, "Ultrathin (<4nm) SiO₂ and Si-O-N gate dielectric layers for silicon microelectronics: Understanding

- the processing, structure, and physical and electrical limits”, *App. Phys. Rev.* Vol. 90 (5), pp. 2057, 2001
- [4:11] R. M. C de Almeida, I. J. R Baumvol, J. J. Ganem, I Trimaille, and S. Rigo, “Thermal growth of silicon oxynitride films on Si: A reaction-diffusion approach”, *J. App. Phys.* Vol. 95(4), pp. 1770, 2004
- [4:12] K. A. Ellis, and R.A. Buhrman, “Nitrous oxide (N_2O) processing for silicon oxynitride gate dielectrics”, *IBM J. RES. DEVELOP.* Vol. 43 (3), pp. 287, 1999
- [4:13] [M.M. Heyns *et al.*, “Cost-effective cleaning and high-quality thin gate oxides”, *IBM J. Res. Develop.*, vol. 43, pp. 339-350, May 1999
- [4:14] M.L. Green, E. P. Gusev, R. Degraeve, and E. L. Garfunkel, “ Ultrathin (<4nm) SiO_2 and Si-O-N gate dielectric layers for silicon microelectronics: Understanding the processing structure, and physical and electrical limits”, *J. Appl. Phys*, vol. 90, pp. 2057-2118, September 2001
- [4:15] M.L. Green, E. P. Gusev, R. Degraeve, and E. L. Garfunkel, “ Ultrathin (<4nm) SiO_2 and Si-O-N gate dielectric layers for silicon microelectronics: Understanding the processing structure, and physical and electrical limits”, *J. Appl. Phys*, vol. 90, pp. 2057-2118, September 2001
- [4:16] Y. Shi, X. Wang, and T.- Ping, “Electrical Properties of High-Quality Ultrathin Nitride/Oxide Stack Dielectrics”, *IEEE Trans. Electron Devices*, vol. 46, pp. 362-368, 1999
- [4:17] Ying Shi, Xiewen Wang, and Tso-Ping, “Electrical Properties of High-Quality Ultrathin Nitride/Oxide Stack Dielectrics”, *IEEE Trans. Electron Devices*, vol. 46, pp. 362-368, 1999
- [4:18] J. W. Harrison and J. R. Hauser, “Theoretical calculations of electron mobility in ternary III=V compounds”, *J. App. Phys.*, Vol. 47(1), pp. 292-300, 1976

-
- [4:19] R.M.C.de Almeida, I. J. R. Baumvol, J. J. Ganem, I. Trimalle, and S.Rigo, "Thermal growth of silicon oxynitride films on Si: A reaction-diffusion approach", *J. App. Phys.*, Vol. 95(4), pp. 1770-1773, 2004
 - [4:20] S. Habermehl, and R. T. Apodaca, "Correlation of charge transport to intrinsic strain in silicon oxynitride and silicon-rich silicon nitride thin films", *Applied physics letter*, Vol. 84(2), pp. 215-217, 2004
 - [4:21] Y. Taur and T. H. Ning, *Fundamentals of VLSI devices*, Cambridge University Press, 1998
 - [4:22] A. Lochtefeld, I. J. Djomehri, G. Samudra, and D. A. Antoniadis, "New insights into carrier transport in n-MOSFETs", *IBM J. RES. & DEV.* Vol. 46, No 2/3, pp. 347-357, 2002
 - [4:23] R. F. Pierret, *Advanced Semiconductor Fundamentals, Volume VI, Modular series on Solid State Devices*, Addison-Wesley Publishing Company, 1987
 - [4:24] E. M. Conwell, and V. F. Weisskopf, "Theory of Impurity Scattering in semiconductors", *Phys. Rev.*, Vol. 77, pp. 388-390, 1950
 - [4:25] S. M Goodnick, R. G. Gann, J.R. Sites, D. K. Ferry, and C. W. Wilmsen, D. Fathy and O. L. Krivanek, "Surface roughness scattering at the Si-SiO₂ interface", *J. Vac. Sci. Technol*, B1(3), pp. 803-808, 1983
 - [4:26] S.Takage, M. Iwase and Toriumi "On the universality of inversion-layer mobility in N- and P-channel MOSFET's" *IEDM Tech. Dig*, pp. 398 - 401, December 1988
 - [4:27] M. Jurczak "Challenges in scaling bulk CMOS technology: ARITMIS results", NESTOR-NanoMOS RIG workshop, Luven, May 2004
 - [4:28] L. Yang, J. R. Watling, F. Adamu-Lema, A. Asenov, and J. R. Barker, "Scaling study of Si and strained Si n-MOSFETs with different high- κ gate stacks", *IEDM Tech. Dig.*, pp. 597 - 600, December 2004.
 - [4:29] G. Baccarani, M. R. Wordman, and R. H. Dennard, "Generalised scaling theory and its application to a 1/4 micrometer MOSFET design", *IEEE Trans. Electron Devices*, Vol. 31, pp. 452, 1984
 - [4:30] R. Hull, *Properties of Crystalline Silicon, emis Data review series, No. 20, an INSPEC publications, The IEE*, London UK, 1999
 - [4:31] R. Chau, J. Kavalieros, B. Roberds, R. Schenker, D. Lionberger, D. Barlage,

- B. Doyle, R. Arghavani, A. Murthy, and G. Dewey, "30nm Physical Gate Length CMOS Transistors with 1.0 ps n-MOS and 1.7 ps p-MOS Gate Delays", *IEDM Tech. Dig.*, pp. 45 - 48, December 2000
- [4.32] H. Wakabayashi, M. Ueki, M. Narihiro, T. Fukai, N. Ikezawa, T. Matsuda, K. Yoshida, K. Takeuchi, Y. Ochiai, T. Mogami, and T. Kunio, "45-nm gate length CMOS technology and beyond using steep halo", *IEDM Tech. Dig.*, pp. 49 - 52, December 2000
- [4.33] K.Goto, Y.Tagawa, H.Ohta, H.Morioka, S.Pidin, Y.Momiyama, H.Kokura, S.Inagaki, N.Tamura, M.Hori, T.Mori, M.Kase, K.Hashimoto, M.Kojima and T.Sugii, "High Performance 25 nm Gate CMOSFETs", *IEDM Tech. Dig.*, pp. 623 - 626, December 2003.
- [4.34] K. Ota, K. Sugihara, H. Sayama, T. Uchida, H. Oda, T. Eimori, H. Morimoto, and Y. Inoue, "Novel locally strained channel technique for high performance 55nm CM", *IEDM Tech. Dig.*, pp. 27 - 30, December 2002.
- [4.35] B. Yu, H. Wang, A. Joshi, Qi , Effiong Ibok, M-R Lin , "15nm Gate Length Planar CMOS Transistor", *IEDM Tech. Dig.*, pp. 937 - 939, December 2001.
- [4.36] K. Ota, K. Sugihara, H. Sayama, T. Uchida, H. Oda, T. Eimori, H. Morimoto, and Y. Inoue, "Novel locally strained channel technique for high performance 55nm CM", *IEDM Tech. Dig.*, pp. 27 - 30, December 2002.
- [4.37] B. Yu, H. Wang, A. Joshi, Q. Xiang, E. Ibok, M.-R Lin, "15nm Gate Length Planar CMOS Transistor", *IEDM Tech. Dig.*, pp. 937 - 939, December 2001
- [4.38] F. Boeuf, T. Skotnicki, S. Monfray, C. Julien, D. Dutartre, J. Martins, P. Mazoyer, R. Palla, B. Tavel, P. Ribot, E. Sondergard, and M. Sanquer, "16nm planar NMOSFET manufacturable within state-of-the-art CMOS process thanks to specific design and optimisation", *IEDM Tech. Dig.*, pp. 637 - 640, December 2001
- [4.39] F. Boeuf, F. Arnaud, B. Tavel, B. Duriez, M. Bidaud, P. Gouraud, C. Chaton, P. Morin, J. Todeschini, M. Jurdit, L. Pain, V. De-Jonghe, M. T. Basso, D. Sotta, F. Wacquant, J. Rosa, R. El-Farhane, S. Jullian, "A conventional 45nm CMOS node low-cost platform for general purpose and low power applications", *IEDM Tech. Dig.*, pp. 425 - 428, December 2004.

5

Intrinsic Parameter Fluctuations in Scaled MOSFETs

In chapter 2 we briefly discussed different sources of intrinsic parameter fluctuations (IPF) and reviewed some of the literature. The IPF are becoming one of the major factors limiting the scaling of conventional MOSFETs and their integration into billion transistor count chips. In this chapter we study IPF due to random dopants and Line Edge Roughness (LER) in well calibrated 35 nm MOSFETs and in the scaled transistors corresponding to the next generations of technology nodes according to the ITRS. Recall from the discussions in chapter 2 that, the main sources of IPF are: random discrete dopants, gate LER and Si/SiO₂ interface roughness leading to oxide thickness fluctuations. However, only the simulation results with the first two sources of fluctuations are presented in this thesis.

Following introductory discussion on discrete random dopants, the results from the statistical simulator on the IPF in scaled conventional MOSFETs due to random discrete dopants are presented in section 5.1. Section 5.2 deals with the simulation of IPF introduced by LER in 35 nm and 25 nm transistors. In the case of the 35 nm device the impact of different LER RMS amplitudes ($\Delta = 3$ nm, 2 nm, 1 nm) is also studied.

In addition to the statistical LER simulations, which are based on randomly generated gate edges, a study of “*deterministic*” LER patterns is presented in section 5.3. In this case the gate edges are intentionally designed to have specific shapes representing specific roughness configurations. A full process simulation with Taurus has been performed, in the case of LER simulation, for both the 35 nm and 25 nm n-MOSFETs.

5.1 Random discrete dopants

The number of dopants in long and wide channel MOSFETs is large and the IPF due to the variations in the number of dopants and their position in the active region are not so significant. However as MOSFETs scale down to sub-100 nanometer dimensions the statistical variation in the number of dopants and their physical location in the active region of devices results in significant IPF. The atomistic nature of these sub-100 nm MOSFETs is illustrated in figure 5:1a [5.1]. The devices, with sub-50 nm dimensions shown in figure 1b and c, can no longer be described, modelled or simulated based on the traditional assumptions of continuous dopant distribution, smooth interfaces and straight gate edges, illustrated in figure 5:1a. The granularity of charge and the atomicity of matter result in structural variations from device to device which is the cause of IPF [5.1] [5.]

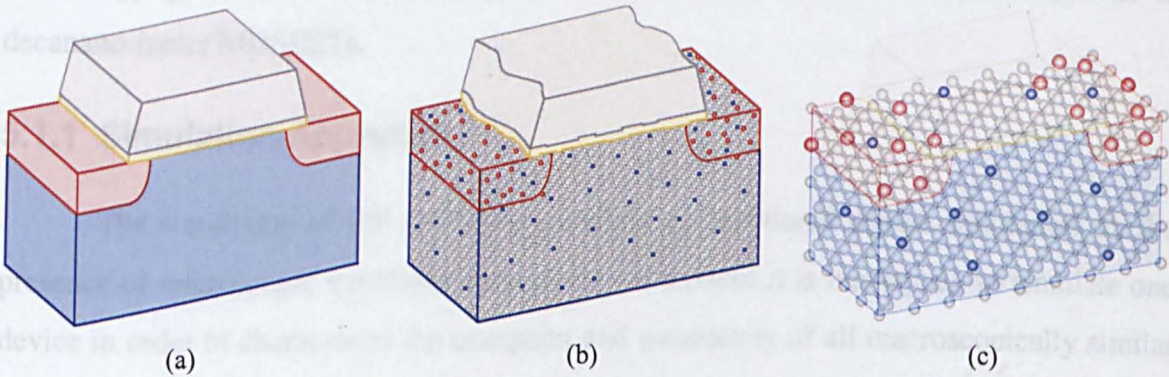


Figure 5:1 The evolution of atomistic devices: (a) MOSFET with continuous ionised dopant charge and smooth boundaries and Si/SiO₂ interfaces; (b) Sketch of 22 nm channel length MOSFET required for 45 nm technology node with rough interface, LER and random discrete dopants; (c) Impression of 5 nm channel length MOSFET with the silicon crystal lattice superimposed. Courtesy of Asenov *et al.*[5:1]

MOSFETs with 7 nm gate lengths are expected to be in mass production around 2018. Such devices, similar to the one in figure 5:1c, will have approximately 10-15 silicon atoms along the channel length and the position of each silicon, dopant or insulator atom is likely to have significant macroscopic impact on the device characteristics. Conventional MOSFETs at such dimensions require a high doping concentration in the channel in order

to suppress short channel effects. Aside from its adverse effect on device performance the high doping concentration will introduce intolerable IPF.

Due to the increasing impact of discrete random dopants on critical device parameters, it becomes very important to consider this phenomenon in the process of nano-scale device modelling and simulation. This means that the introduction of statistical analysis into process and device simulation is vital. The Glasgow University atomistic device simulation software (discussed in chapter 3) is a specifically developed simulation tool to investigate this problem, which is one the major stumbling blocks of CMOS device scaling.

Most simulation studies of intrinsic parameter fluctuations have been restricted to devices corresponding to one particular technology node, or to idealized devices. In this work, we have extended the study of intrinsic parameter fluctuations to a family of realistic MOSFETs corresponding to all technology nodes until the end of the current edition of the ITRS. This allows us to investigate the magnitude of parameter fluctuations in future technology generations and will enable us to understand the trend of fluctuations in decanano-meter MOSFETs.

5.1.1 Simulation Approach

The simulation of IPF shifts the paradigm of traditional device simulation. In the presence of microscopic variations from device to devices it is inadequate to simulate one device in order to characterize the operation and parameters of all macroscopically similar but microscopically different devices. It becomes necessary to simulate a statistically significant sample of devices. This will allow a meaningful statistical analysis including estimates of mean values, variances and higher moments of the statistical distributions.

In this section, using statistical 3D simulations, we study the random dopant induced IPF in conventional MOSFETs scaled to 25, 18, 13, and 9 nm gate lengths according to the requirements of the ITRS for high performance devices in the 65, 45 32 and 22 nm technology nodes respectively.

In the scaling procedures, described in more details in chapter 2, the generalized scaling rules, in parallel with ITRS predictions' have been adopted. The doping concentration was broadly scaled by the factors, α and κ ($N'_a = \alpha\kappa N_a$), where $\kappa \approx \sqrt{2}$ and $\alpha = \kappa(V'_{dd}/V_{dd})$ where, V'_{dd} and N'_a are supply voltage and channel doping concentration

of the scaled devices respectively. Super-retrograde channels have been achieved using ion implantation. The doping in the middle of the channel is reduced by progressive use of pocket implants (see chapters 3 and 4).

The assignment of discrete dopants is based on rejection technique [5:2]. Given the expected number of dopants estimated from the continuous doping distribution obtained from the Taurus process simulator [5.3], the probability that there is a dopant in a mesh ‘brick’ (a cell of 3-D mesh) is calculated. Then, using rejection, the dopants are placed randomly according to the initial continuous-doping distribution.

During the atomistic device simulation mobility is handled as follows. Based on the results of calibration and scaling obtained from full process simulation steps, a continuously doped device is simulated for all I - V points and the corresponding mobility profile is stored. This mobility profile is then used in the course of the atomistic device simulation

A typical simulation domain used in the simulation of the 35 nm Toshiba MOSFET is illustrated in figure 5:2. The position of individual discrete dopants is clearly identifiable from the potential landscape. The fluctuation in the surface potential and depletion layer edge inflicted by individual dopants are clearly seen.

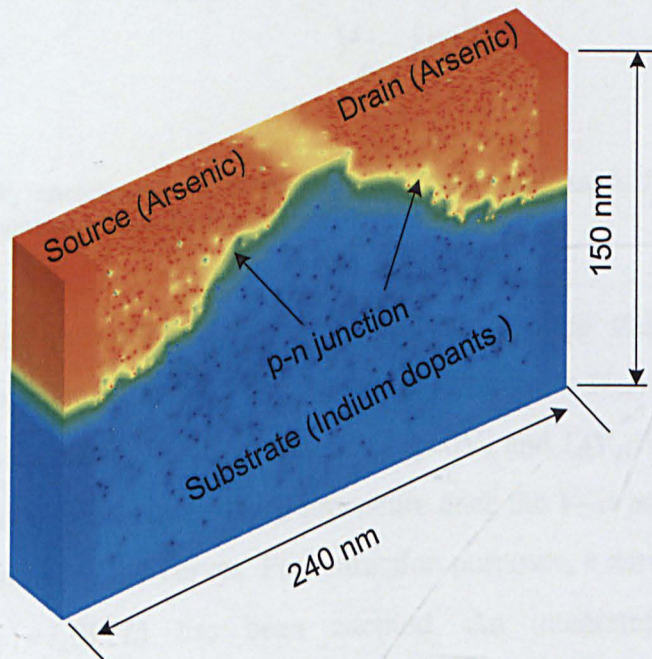


Figure 5:2 Image of 3D potential distribution in one of the simulated 35 nm MOSFETs illustrating the position of discrete random dopants in the channel (indium) and source/drain regions (arsenic). The continuous yellow line depicts the approximate position of p - n junction.

The spread of the current voltage characteristics of a sample of 200 microscopically different devices is illustrated in figure 5:3. The open-dotted line represents an experimental data from [5.4]. Results of the statistical analysis of such devices are reported in the next subsections.

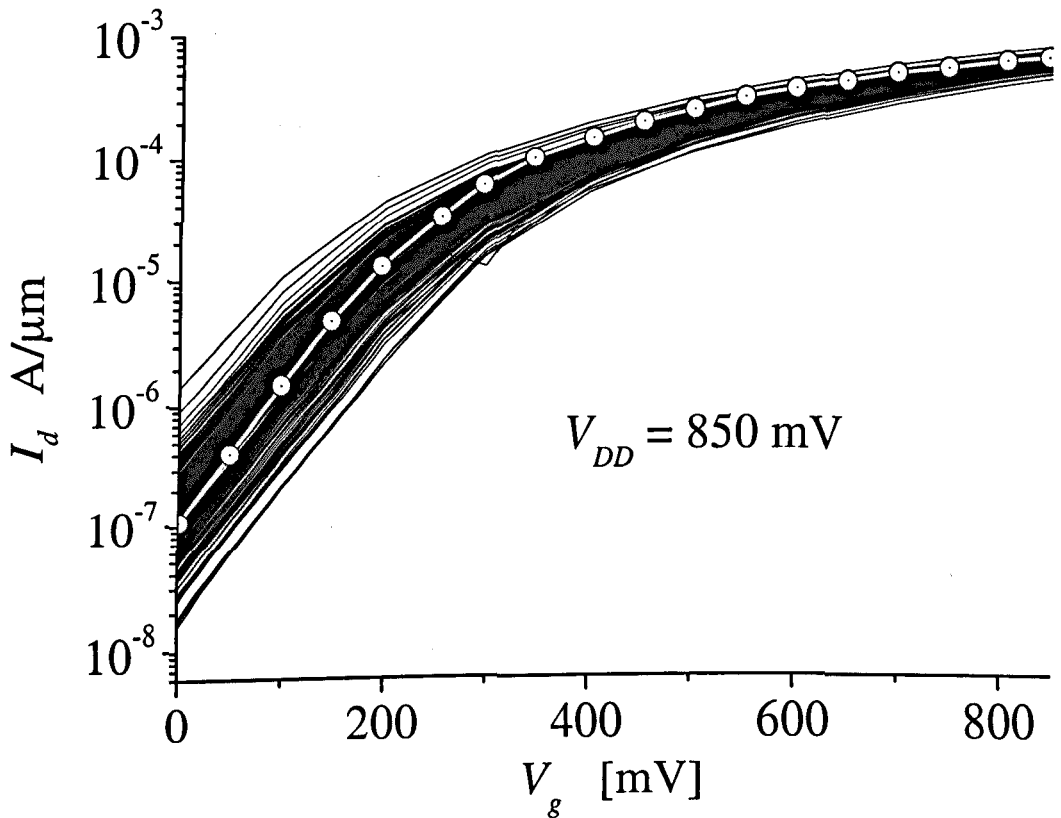


Figure 5:3 Spread I_d - V_g characteristics of 200 transistors with 35 nm gate length. The open circle shows the I - V curve from the experimental 35nm Toshiba device.

Three device parameters namely, V_T , I_{inv} and I_{off} are statistically studied and presented in the following sections. The values of off-current (I_{off}) and strong inversion current (I_{inv}) are extracted at the points where $I_d(V_g=0V)$ and $I_d(V_g=V_{dd})$ respectively. σV_T was calculated using a standard statistical procedure once the V_T is automatically extracted from each of the simulated I - V curves. For extraction purposes, a current threshold voltage criterion $I_d(V_g \approx V_T) = I_o(W/L)$ has been adopted. An integrated device simulation methodology, described in [5.5], has been implemented. In the following sub-sections the results of statistical simulation of IPF due to the number of dopants and their position in the depletion region of the scaled MOSFETs is presented.

5.1.2 Threshold Voltage variation

Threshold voltage is an important parameter in the MOSFET design. In concert with the subthreshold slope it determines the off-state leakage current [5.6]. The gate voltage overdrive, which is the difference between the supply voltage and the threshold voltage determines the drive current. One of the main roles of the channel profile design is for V_T adjustments [5.7] in conjunction with the short channel control. Well defined, steady and stable threshold voltage a crucially important for analogue and digital circuits, *i.e.*, less variation of V_T is highly desirable.

Therefore it is, important to keep σV_T within an acceptable degree of tolerance in order to deliver a stable integrated circuit and properly working system. However in real nano-scale MOSFETs and integrated circuits the discrete random dopants in the channel region introduce V_T fluctuations. Moreover, the fluctuations increase significantly as the gate length decreases

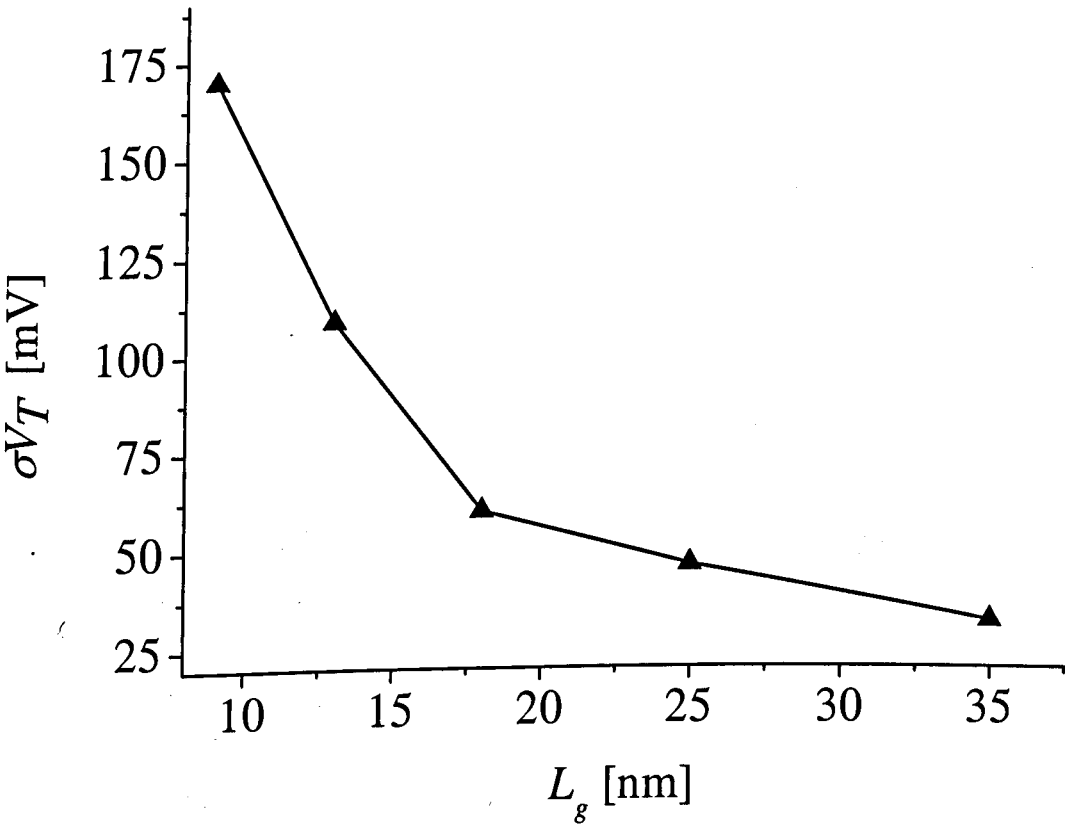


Figure 5:4 Threshold voltage variation as a function of scaled device gate length.

Figure 5:4 shows the standard deviation of the threshold voltage, σV_T , as a function of the gate length shrinks. σV_T rapidly increases when the transistors are scaled below 20 nm gate length. For example a standard deviation of approximately 60mV for the 18 nm transistor gives a spread in $\pm 3\sigma V_T$ of $\sim \pm 360$ mV, which is very large, resulting in some of the devices in the integrated circuit not turning on. As the channel length is further reduced the fluctuations become excessively high with a spread for $\pm 3\sigma V_T$ of $\sim \pm 510$ mV for the 9 nm transistor. This is in excess of the supply voltage expected to be in the range of ($V_{dd} \sim 500$ -600mV).

Although most studies focus on conventional MOSFETs, the V_T variation due to random discrete dopants is not a unique phenomenon for these devices. A recent simulation study in [5:8] showed that the random discrete dopants in source and drain of double gate MOSFETs also induce V_T fluctuations. For example the double gate MOSFET with effective channel lengths of 4, 6, 8, 10 nm exhibits a $\pm 3\sigma$ variation of 204, 90, 45, 30mV respectively. Although the degree of σV_T in the conventional MOSFETs presented in this thesis is much higher, the fluctuations in non-conventional MOSFETs are not negligible [5.8].

The histogram distributions of the threshold voltage in the prototype 35 nm and scaled 25, 18 nm transistors are illustrated in figure 5:5. All the results are on 3D atomistic device simulation of samples of 200 devices each. For all these sets of devices the distribution of the threshold voltage is close to the normal distribution, with a standard deviation increasing with reduction of the channel length.

Another important information that can be drawn from frequency distribution of threshold voltage in the all three devices is the increase in range of V_T distribution as the transistor channel length decreases. The ranges of distribution are approximately 162, 249, and 330mV for 35, 25 and 18 nm gate lengths respectively (see figure 5:5). Comparing between the range of V_T distribution of the 25 nm transistor and the 35 nm prototype transistor, there is a 52% increase and in the case of 18 nm device the range increases more than 200%. This reflects an increase in V_T fluctuation in the smaller channel length transistors.

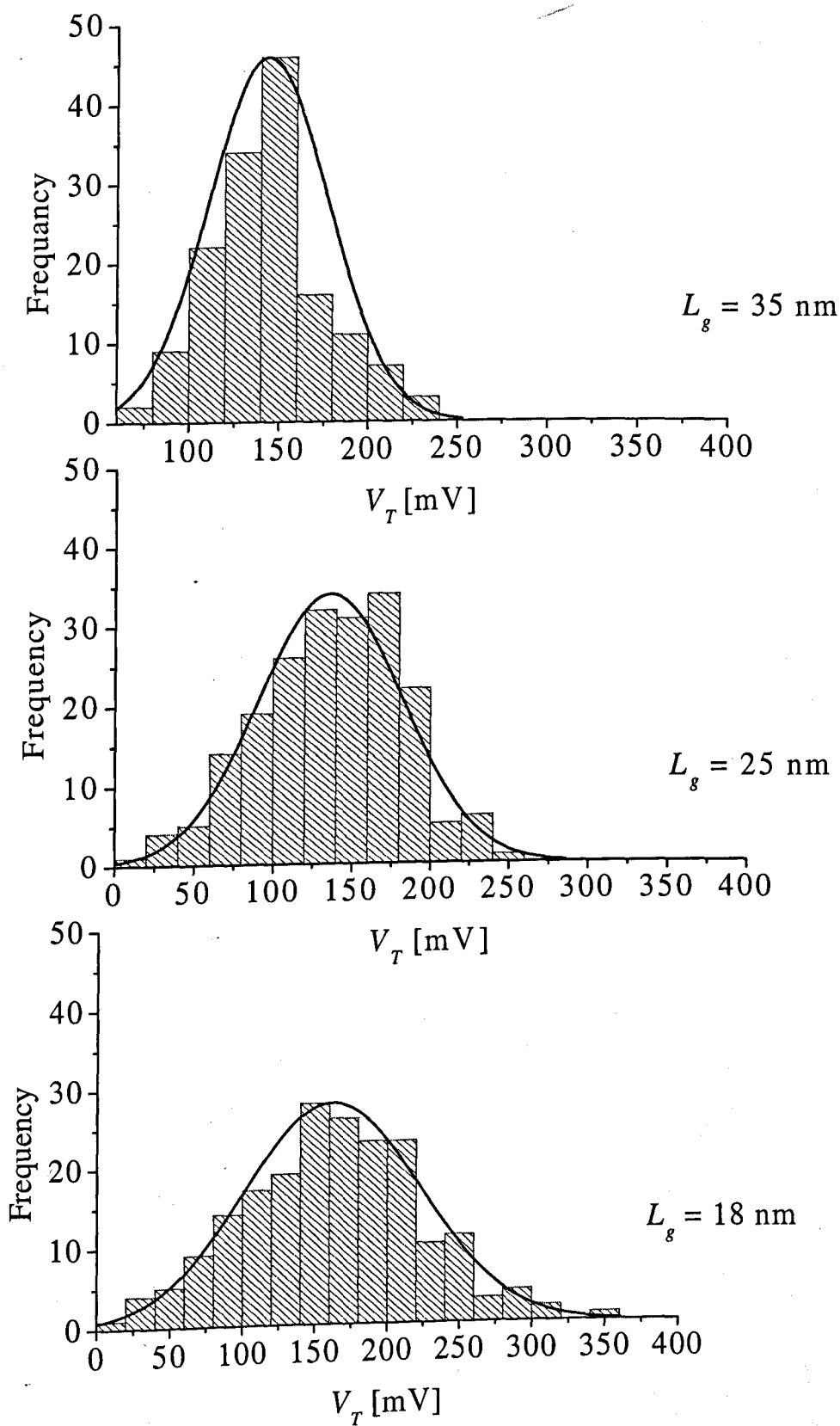


Figure 5.5 Histograms of V_T distributions for the prototype 35 nm and the scaled 25 and 18nm MOSFETs

5.1.3 Off-current variation

The current flow in a MOSFET has two major components: the diffusion current resulting from the gradient carrier concentration and drift associated with the applied electric field [5.7]. While the strong inversion regime is dominated by the drift component of the overall drain current, the subthreshold current is dominated by diffusion [5.9] [5.10].

The *off*-current is another important parameter in MOSFETs, which to great extent determines the standby leakage current in integrated circuits. The impact of discrete random dopants on I_{off} is well illustrated in figure 5:3. The off-current fluctuations are in the range of $1.2 \times 10^{-8} - 1.5 \times 10^{-6}$ A/ μ m. There are nearly two orders of magnitude differences between the maximum and minimum subthreshold currents in the sample. Such a level of fluctuations is highly unacceptable for particular circuit applications, like 6T SRAM cells (see section 5.1.5).

The mean and the standard deviation of I_{off} is depicted in figure 5:6. The presence of random discrete dopants results in a substantial increase in the average subthreshold leakage current with the reduction of a channel length. For example σI_{off} for 18 nm is 220nA/ μ m, which is more than 180% greater than the mean off current for a 35 nm MOSFET. This is highly undesirable from the static power dissipation point of view.

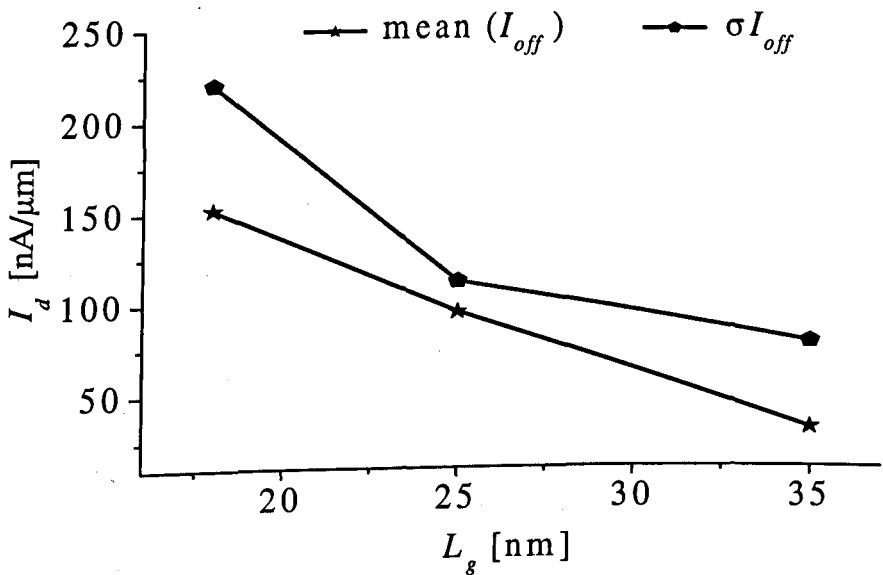


Figure 5:6 Mean and standard deviation of I_{off} in the 35 25 and 18 nm MOSFETs

In order to visualize and compare the distributions the *off*-current for different channel lengths, the histograms of these distributions are presented for all three devices in figures 5:7. The column one represent plot of linear scale and two represent $\log(I_{off})$

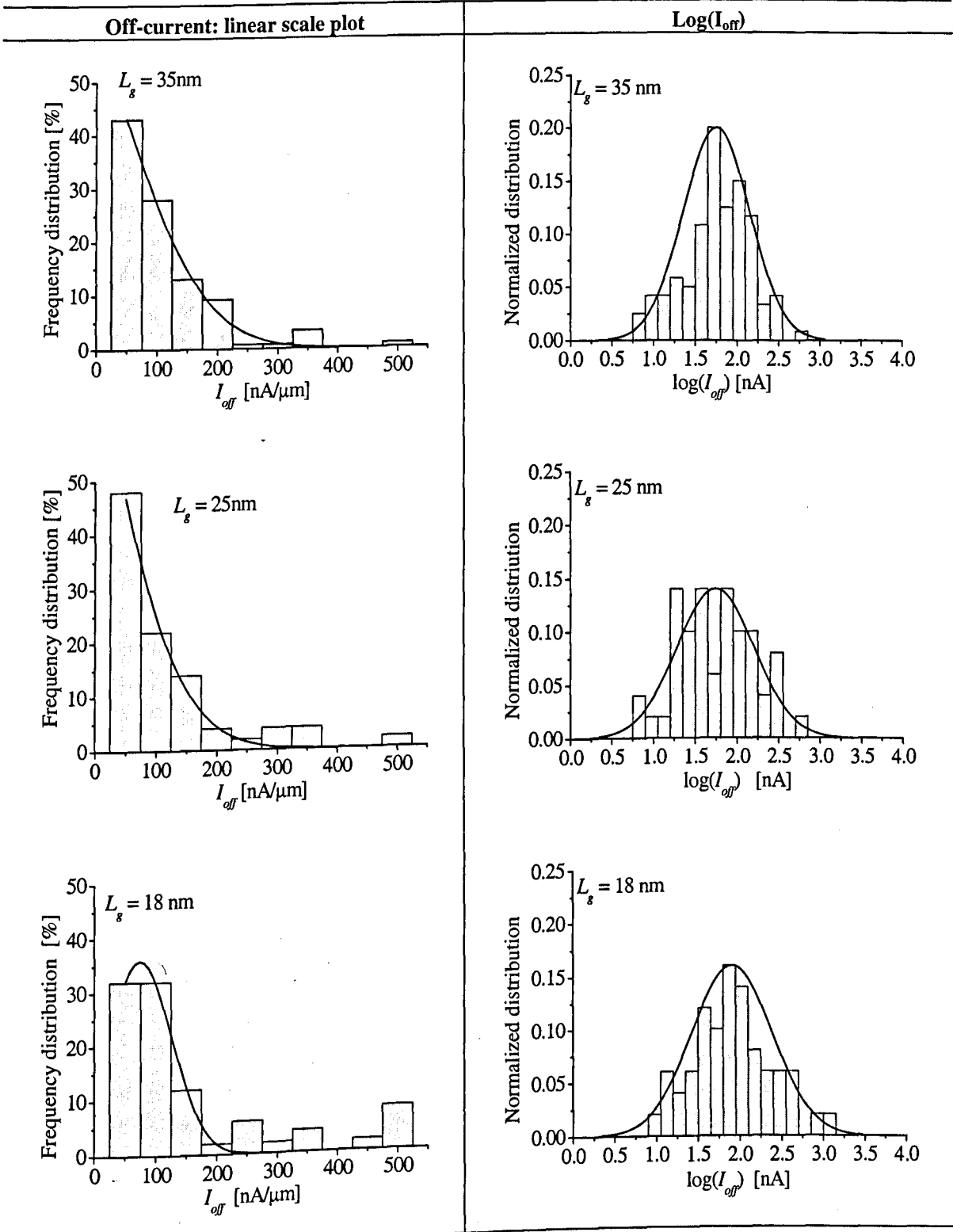


Figure 5:7 I_{off} distribution of the calibrated and scaled (25 and 18nm) devices both in linear scale and logarithmic scale.

Unlike the distributions of the threshold voltage, which are close to normal distributions, the I_{off} distribution is a log-normal distribution (see figure 5:7 above). This is due to the exponential dependence of the subthreshold current on gate voltage. It is also important to note that there is substantial increase in the average subthreshold leakage current with the reduction of channel length. For example the average I_{off} in 25 and 18 nm device samples are 95 and 152 nA/ μ m respectively. This corresponds to a percentage increase of 11.7% and 72.7% respectively compared to the off-current in the 35 nm transistor.

Figure 5:8 illustrates the channel length dependence of the extracted skew and kurtosis of the distributions of the off-current which in this case differ substantially from the expected value for the normal distribution. The other important information that can be drawn from the results plotted on figure 5:8 is that the smaller the gate length of the devices the more the distribution is skewed to the right (*positive skew*) with reference to the prototype device of 35 nm gate length.

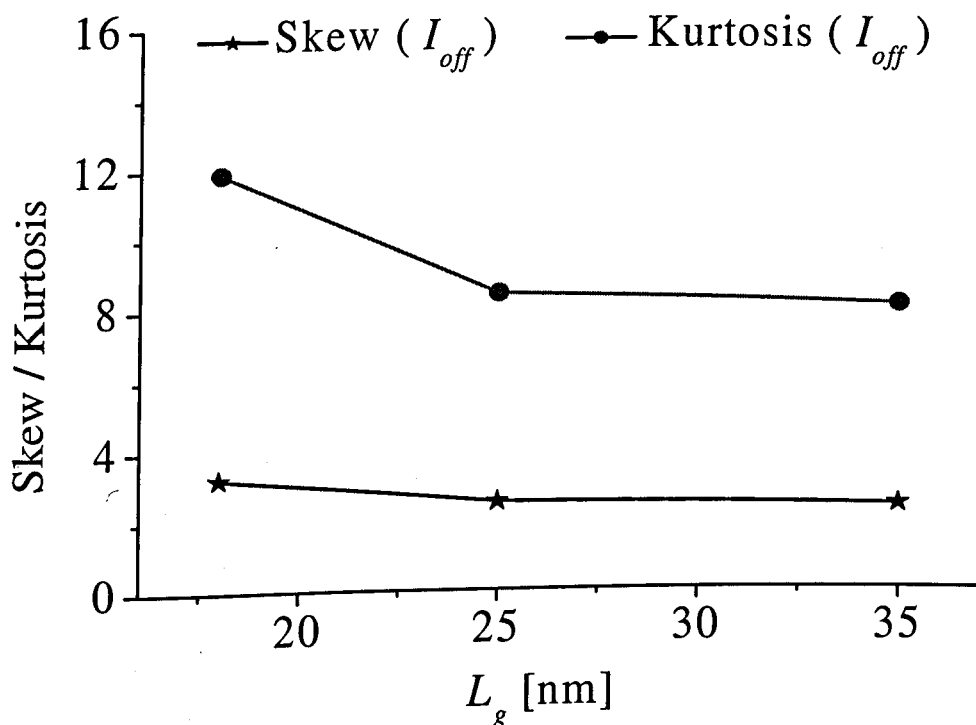


Figure 5:8 3rd and 4th moments of the I_{off} distribution in 18, 25, and 35nm MOSFETs

5.1.4 Drain current variation in the strong inversion region

For simplicity and computational efficiency, we have studied the drain current in strong inversion at low drain voltage as an indication of the fluctuations in the drive current (I_{on}). Although there will be a *quantitative* difference in the I - V simulation results under high drain voltage, the *qualitative* difference estimated in previous simulation studies [6.2] is rather small. Hence the results presented in this section are good indications to the fluctuation of “ I_{on} ” despite the fact that the data are obtained from the atomistic device simulation at low drain voltage of 50mV.

The corresponding data are extracted in strong inversion at gate voltage of $V_g = 850\text{mV}$ which is equivalent to the supply voltage of $V_{dd} = 850\text{mV}$. The drain current fluctuations in strong inversion are not as strong as in the subthreshold current. At high gate voltage the inversion layer charge screens the Coulomb potential of the individual discrete dopants reducing the potential fluctuations in the channel and the corresponding current fluctuations.

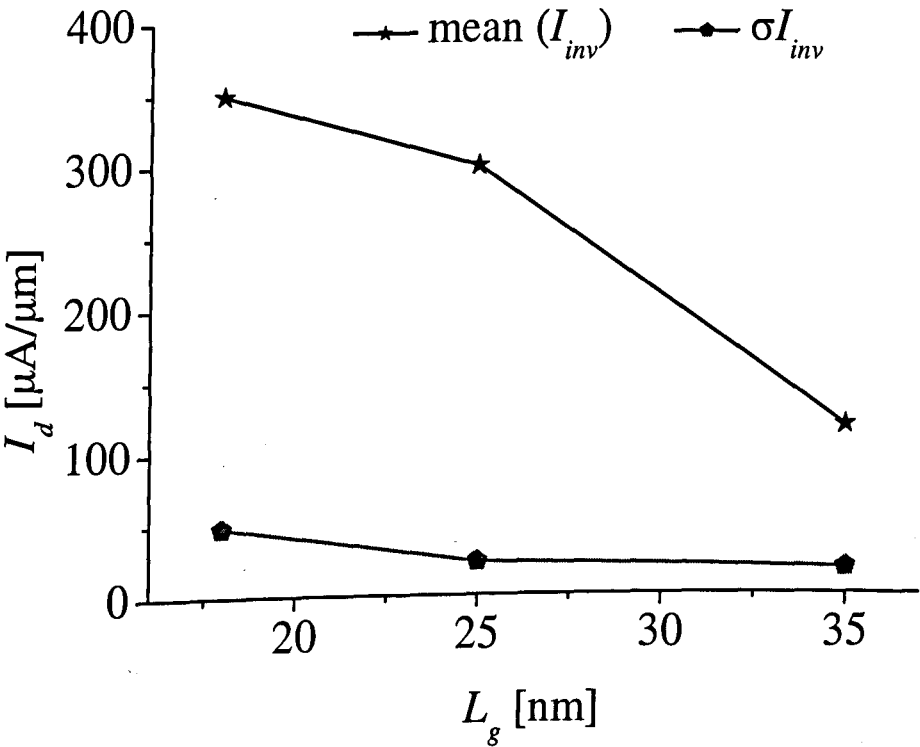


Figure 5:9 Standard deviation and mean of drain current in the strong inversion regime.

As expected there is a trend of increasing I_{on} fluctuation magnitude with the reduction of the channel length illustrated in figure 5:9. For example, the $\pm 3\sigma$ variation of the strong inversion current for an 18nm transistor is about $288\mu\text{A}/\mu\text{m}$, which is a significant current variation in such a small device. As with the distribution of the threshold voltage, the distribution of the on-current is also close to normal in form (see figure 5:10).

The inversion drain current is crucial to the performance of MOSFETS. One of the motivations of scaling down transistors size is to achieve high relative drive current in order to boost circuit performance. However variation in the drive current may result in mismatch in analogue applications and variation in the signal propagation times of digital circuits.

The mean and the standard deviation of I_{inv} and I_{off} for the different channel lengths are summarised in table 5:1. The skew and kurtosis of the off-current distribution is also added to the table. The I_{inv} current variations are relatively smaller comparing it with the corresponding I_{off} variations. On the other hand, the increases in the fluctuation of I_{inv} in the scaled devices are not negligible. There is a 35% increase of σI_{inv} in the 25 nm transistor compared to the prototype 35 nm transistor. This percentage of fluctuation in I_{inv} increases to 176% in the 18 nm transistor.

	Average		Standard deviation			Skew	Kurtosis
L_g [nm]	$\langle I_{inv} \rangle$ [$\mu\text{A}/\mu\text{m}$]	$\langle I_{off} \rangle$ [nA/ μm]	σI_{inv} [$\mu\text{A}/\mu\text{m}$]	σI_{off} [nA/ μm]	σV_T [mV]	I_{off}	
35	282	35	18	78	33	2.4	8.1
25	300	95	23	112	46	2.5	8.5
18	350	152	48	220	61	3.0	12.0

Table 5:1 Summary of selected statistical parameters

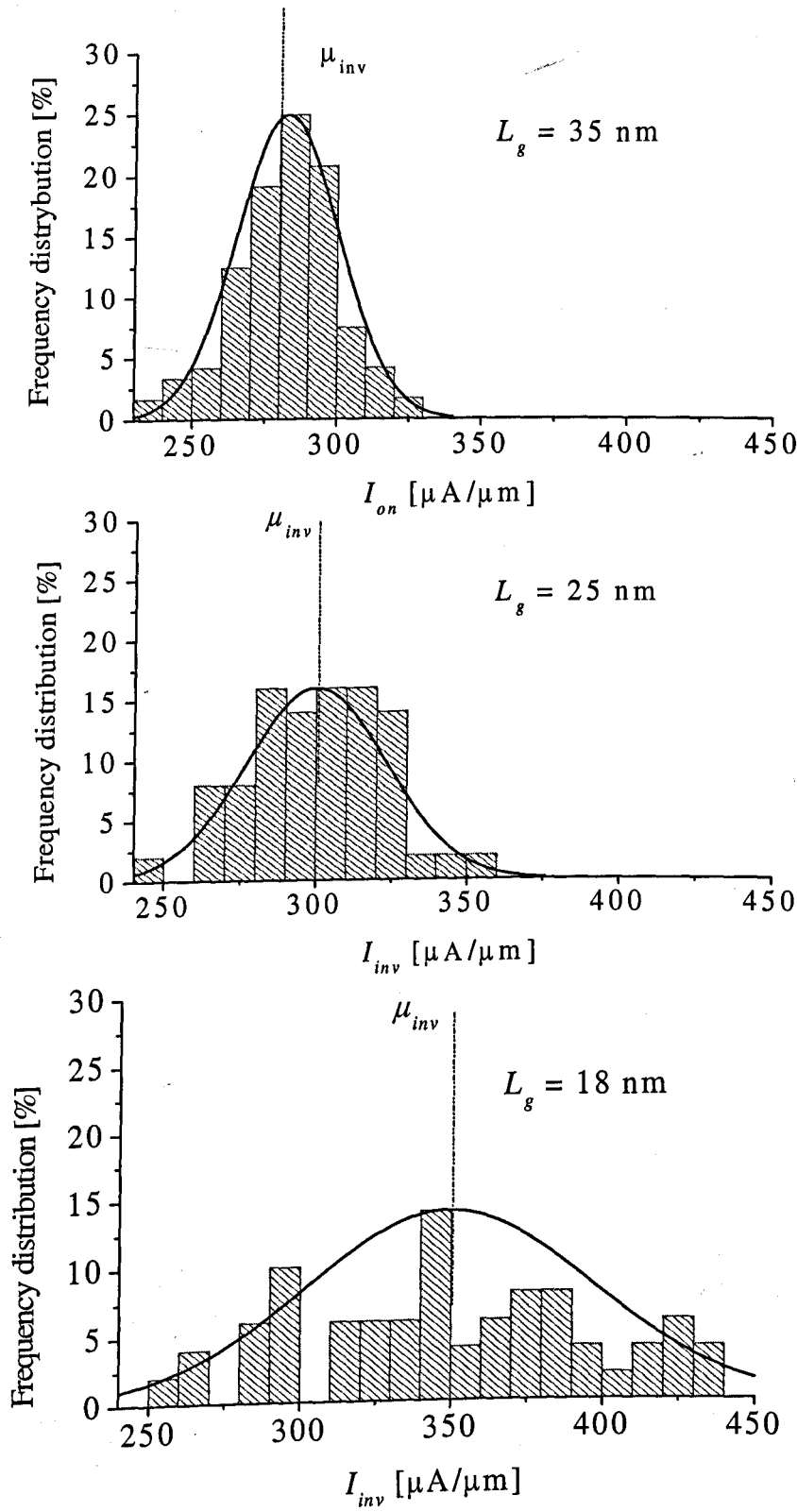


Figure 5:10 Normalized histogram of drain current distribution in the strong inversion region (I_{inv}) of the 35 nm; the 25 nm and 18 nm n-channel transistors.

5.1.5 Impact of random discrete dopant fluctuation on circuits

The IPF resulting from random dopants in individual devices discussed in the previous sections contribute significantly to mismatch [5.11] between neighboring devices in integrated circuits. Recent simulation studies show that random doping fluctuations have already adversely affected the yield and the stability of SRAM illustrated in figure 5:11 [5.12][5.13].

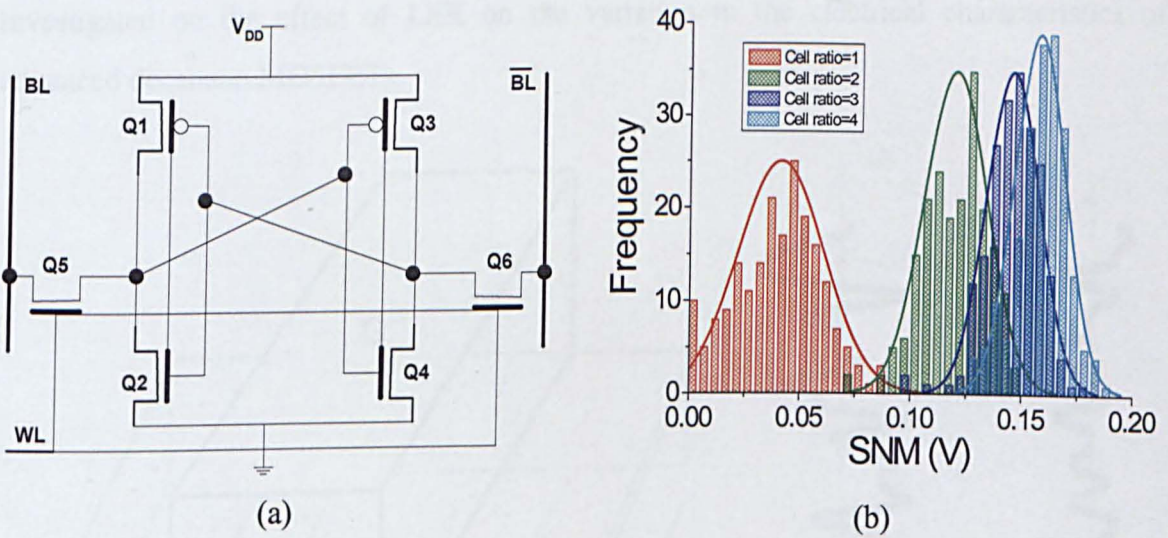


Figure 5:11 Circuit schematics of SRAM cell (a); Distribution of SNM; adapted from [5.14]

Figure 5:11 illustrates the random dopant induced distribution of static noise margin (SNM) in an ensemble of SRAM cells of various cell ratios at the transition characteristic for the advanced stages of the 90nm technology node. To obtain acceptable yield, in the presence of random dopants fluctuation, the *cell ratio*, $r = (W_d / L_d) / (W_a / L_a)$ (where W and L are width and length of the driver and the access transistors in SRAM cell) must be increased from its nominal value of $r = 1$ to $r = 3$. However, this increase in cell ratio leads to a larger cell area and thus compromises SRAM scaling and hence the scaling of the majority of silicon based digital systems. Moreover, the simulation study in [5:14] also shows that fluctuation of SNM due to random dopants in the 35 nm device is comparable with the fluctuation caused by instabilities in the power supply of $\pm 25\%$.

5.2 Line edge roughness in conventional MOSFETs

Ideal MOSFETs are considered to have a straight gate edges as illustrated in figure 5:10a. However in real devices, the gate is prone to line edge roughness (LER) inherited from various imperfect process steps. In the past, LER had little impact on device operation because the gate length was much larger compared to the roughness of the gate edges. However, with transistors gate length scaled to sub-50nm the contribution of LER to overall IPF is becoming proportionally significant. Therefore, in this work, we have investigated on the effect of LER on the variation in the electrical characteristics of advanced decanano MOSFETs.

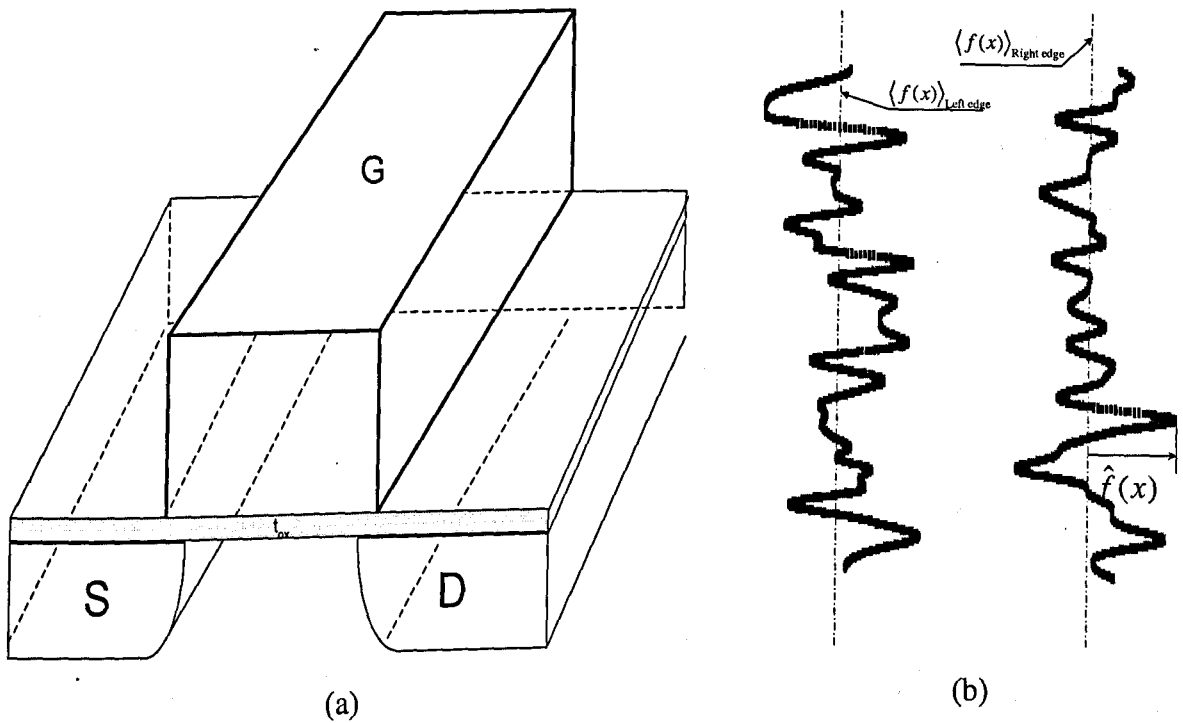


Figure 5:12 Comparative illustrations of (a) an ideal transistor with straight gate edges and uniform channel width and (b) a top view of hypothetical gate edges which are constructed from a pair of the randomly generated rough lines used in this work for to investigate LER

In the following section, brief discussions of the main sources of LER are followed by results from process and device simulations that take in to account the effect of LER on the transistor's parameters. The simulation results are for gate lengths 35 and 25 nm.

5.2.1 Contributing factors to LER

LER is a major process related problem, which affects MOSFET parameters. As a result of gate LER, gate geometry will vary from transistor to transistor. Since the drain current is related to the gate geometry the overall current may vary introducing also V_T variation from transistor to transistor [5.15]. At the same time, the doping distribution near the p-n junctions will follow the gate shape introducing variations in the channel gate length. It is therefore important to control the LER in order to minimize the corresponding IPF.

The main contributing factors for LER are: molecular structure (polymer aggregate) of resist,* which will generate a line width fluctuation in resist pattern that will affect the final shape of the transistor line edge during the etching process [5.16]; the degree of mixing of resist materials (photosensitive compound, base resin, and organic solvent); the fluctuation of photo chemical events, like absorption of radiation and reflected photons and the pattern transfer process (etching) where the acid diffusion length can degrade the line edge and space pattern [5.17].

5.2.2 Characterisation of LER

LER can quantitatively be characterized by an autocorrelation function with two parameters; a correlation length (Λ) and the root mean value (Δ). From a statistical point of view, the autocorrelation function (ACF) describes the dependence of values of random data at one interval on the values at another one [5.18]. It can directly be estimated from the sampled data values as the product of a distance dx -shifted with itself [5.19] [5.20]. Assuming that $\langle f(x) \rangle = 0$ over a large sample of discrete data, the ACF can be expressed as:

$$ACF(r\Delta x) = \frac{1}{N-r} \sum_{n=1}^{N-r} (f(x_n))(f(x_{n+r})), \quad (6.1)$$

$$r=1,2,\dots,m: (m < N)$$

* The role of resist materials in the device manufacturing process is to perform the function of transforming a pattern placed on the mask onto a desired location.

in the case of LER $f(x_n)$ is an estimation to the displacement from the average line $\langle f(x) \rangle$ at point x_n (see figure 5:12b), N is the number of samples regularly spaced by Δx over the width of the transistor, r is often called the *lag number* and m is its maximum value [5.21]. Usually LER is assumed to have a Gaussian or exponential $ACF(x)$ given by equations (5.2) and (5.3) respectively.

$$ACF_G(x) = \Delta^2 \exp\left(-x^2 / \Lambda^2\right) \quad (5.2)$$

$$ACF_E(x) = \Delta^2 \exp\left(-x / \Lambda\right) \quad (5.3)$$

There is an additional parameter used for LER characterization known as a roughness exponent (α) discussed in the literature [5.20], [5.22], [5.23], [5.24] that takes into account the relative contribution of high frequency fluctuation to the spatial roughness. It is usually associated with LER in *photo-resists* and 2D-rough surfaces (for example Si/SiO₂ interface roughness). However, in this work we concentrate on the two parameters, Δ and Λ of the correlation function, which we believe are sufficiently adequate to characterize the gate LER.

The parameter Δ characterizes the deviation of roughness magnitude with respect to the stationary length $\langle f(x) \rangle$ (see figure 5:12b). The equation for Δ has a similar definition as the standard deviation in statistics and is given by:

$$\Delta = \sqrt{\sum_{i=1}^N \frac{(f(x_i) - \langle f(x_i) \rangle)^2}{N}} \quad (5.4)$$

Despite the simplicity of equation 5.4 to measure the vertical roughness (Δ) it does not capture the property of roughness along the gate width [5.22]. Hence additional parameter to characterize LER is essential.

The ITRS (2001) defines LER as 3Δ . However in the latest ITRS 2003 edition a new definition has been introduced. The new lithography guide line for LER is defined in terms of line width roughness (LWR) as $LER = \frac{LWR}{\sqrt{2}}$, and the required figures are given in table 6:2. The work on LER in this thesis has been performed prior to the guidance line; hence the values used in this work do not exactly parallel the roadmap. However the result obtained still highlights the problem that the presence of LER creates for devices.

Year of production	2004	2007	2010	2013	2016
Technology nodes	<i>hp90</i>	<i>hp65</i>	<i>hp45</i>	<i>hp32</i>	<i>hp22</i>
Dram $\frac{1}{2}$ pitch	90	65	45	32	22
Printed gate length [nm]	53	35	25	18	13
Printed gate CD* control (3σ) [nm]	3.3	2.2	1.6	1.2	0.8
Line Width roughness (3σ) [nm]	3.0	2.0	1.4	1.0	0.7

Table 5:2 ITRS guidelines for LER, LWR, and printed gate critical dimension (CD) variation

The other parameter needed to more completely characterize the gate LER is Λ , which is a measure of the damping distance or the width of the autocorrelation function. Λ signifies the narrowness of the power spectrum [5.25] of the roughness along the gate width. The value of Λ often used is in the range of 20-50 nm. For most work presented in this thesis, the correlation length is taken to be 30 nm.

5.2.3 LER Simulation approach

In this work the investigation of LER and its effect on transistor parameters is mainly based on full device process simulation, in which the randomly generated rough line pattern is introduced onto the nominally straight gate in order to produce an LER effect during device simulation. The random lines are generated using a Fourier technique explained in detail in [5.26].

* Critical dimensions (CD) are referred to dimensions of the smallest geometrical features like printed gate length, width of interconnect line, contacts, trenches, etc., which can be formed during device manufacturing process. [5.27]

The RMS height and the correlation length are used as inputs to the random rough line generator along with the corresponding choice of autocorrelation function, either the Gaussian or exponential ACF. Various values of RMS amplitude ($\Delta = 1$ nm, 2 nm, and 3 nm) and a correlation length of $\Lambda = 30$ nm were used for each set of process simulated devices. In addition to this further values of Λ (5 nm and 15 nm) were considered to study how the gate shape and the metallurgic p-n junctions are affected by the autocorrelation length.

The randomly generated rough lines of both gate edges are then used as a mask to transfer the roughness pattern to the poly gate during the etching process simulation step. The device structure obtained from the full process simulation is used as input to device simulation. In order to identify the contribution of LER to IPF, all the device process conditions including doping concentration and annealing time are identical to the originally calibrated and scaled devices with the exception of the introduction of LER to the poly gate.

5.2.4 IPF due to LER

In this section the effects of LER on the electrical parameters of MOSFETs including I_{off} , I_{on} and V_T are presented for different Δ and Λ .

5.2.4.1 The effect of LER on device parameters with various value of Δ

Figure 5:13 shows the simulated I_d - V_g characteristics of 50 n-MOSFETs, with nominal gate length of 35 nm, gate width 100 nm and a randomly generated gate edges. Values of Δ and Λ are 1 nm and 30 nm respectively. The inset images show the 2-D net doping profile of the Si/SiO₂ interface of two extreme cases from the sample corresponding to I_{max} and I_{min} . The two devices are very different in terms of effective channel length. Compared to the effective channel of the 35nm MOSFET, which is 33 nm, 'device A' and 'device B' have 22% smaller and 12% larger channel length respectively.

The approximate position of the polysilicon gate is also indicated in the figure 5:13. The p-n junction follows the shape of the gate. In order to study in more detail the effect of LER on impurity doping distribution and carrier mobility in those two extreme cases, the doping profile at the cross-section C-C is shown in figures 5:14 and 5:15.

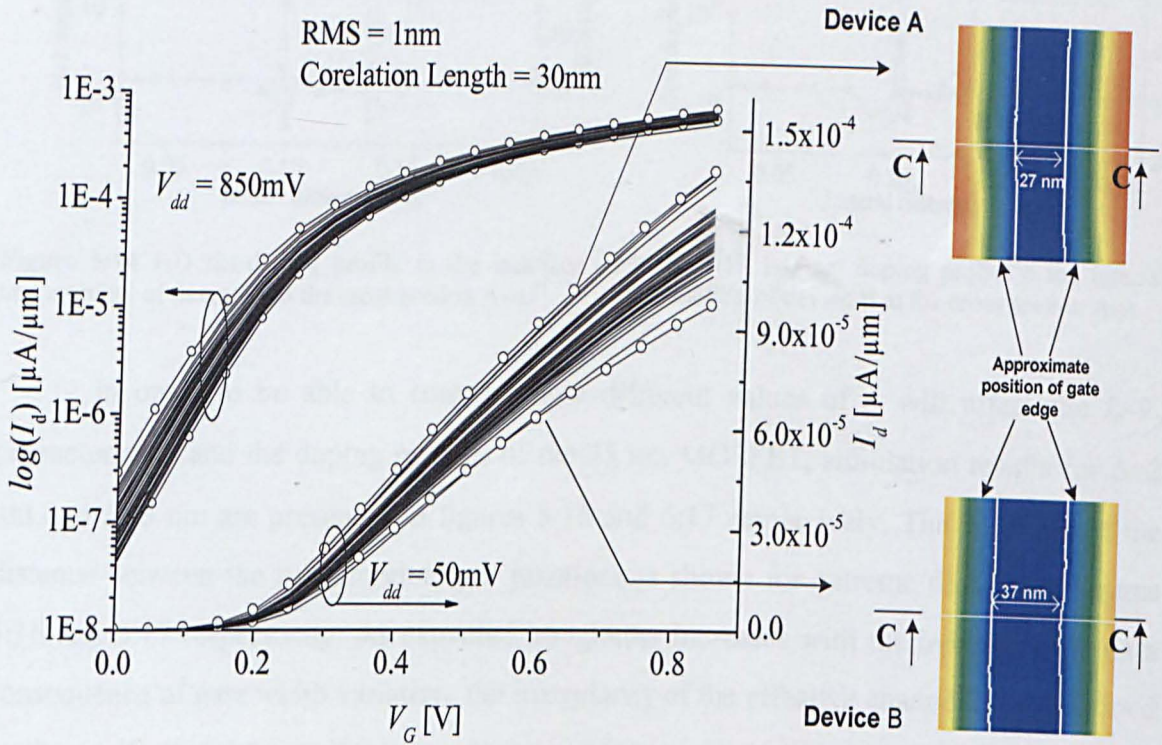


Figure 5:13 I_d - V_g characteristics for a sample of 50 n-channel MOSFETs, including LER effect with an RMS height of 1 nm. Values at high drain voltage of 850mV are plotted in logarithmic scale and those at low drain voltage of 50mV are plotted in linear scale. The image to the right is the net doping profile of the two samples with the minimum and maximum off currents.

The main reason for the low performance of device B and the high performance of device A are the large and the small effective channel length respectively. At the same time according to figure 5:14, device A has lower doping concentration in the channel and higher average channel mobility, $\langle \mu_{ch} \rangle = 60.7 \text{ cm}^2 / Vs$, compared to device B. In addition to larger effective channel length, device B has higher doping concentration in the channel giving average channel mobility, $\langle \mu_{ch} \rangle = 51.5 \text{ cm}^2 / Vs$. Hence device B delivers lower performance.

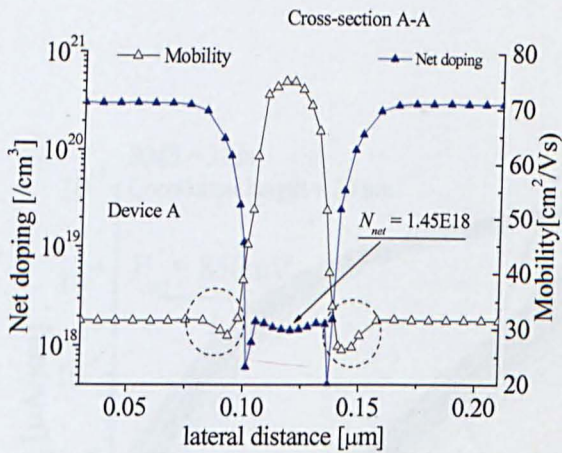


Figure 5:14 1-D net-doping profile in the interface and mobility of device A at the cross section A-A

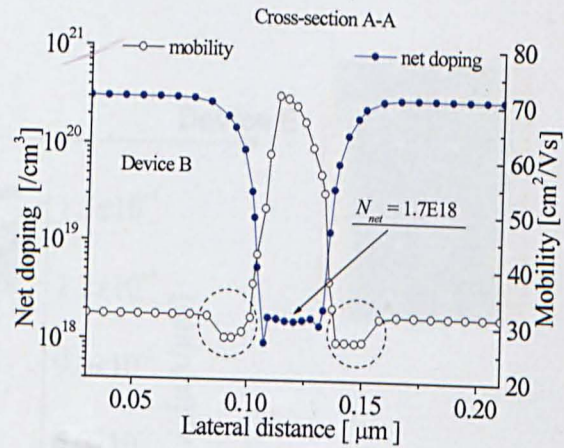


Figure 5:15 1-D net doping profile in the interface and mobility of device B at the cross section A-A

In order to be able to compare how different values of Δ will affect the I_d - V_g characteristics and the doping profiles of the 35 nm MOSFET, simulation results for $\Delta=2$ nm and $\Delta=3$ nm are presented in figures 5:16 and 5:17 respectively. The variation of the distance between the metallurgical p-n junctions is shown for extreme devices in figures 5:16 and 5:17 respectively. As expected, roughness increases with the increasing Δ . As a consequence of gate width variation, the irregularity of the effective channel length defined by the position of the metallurgical p-n junction of the transistors increases.

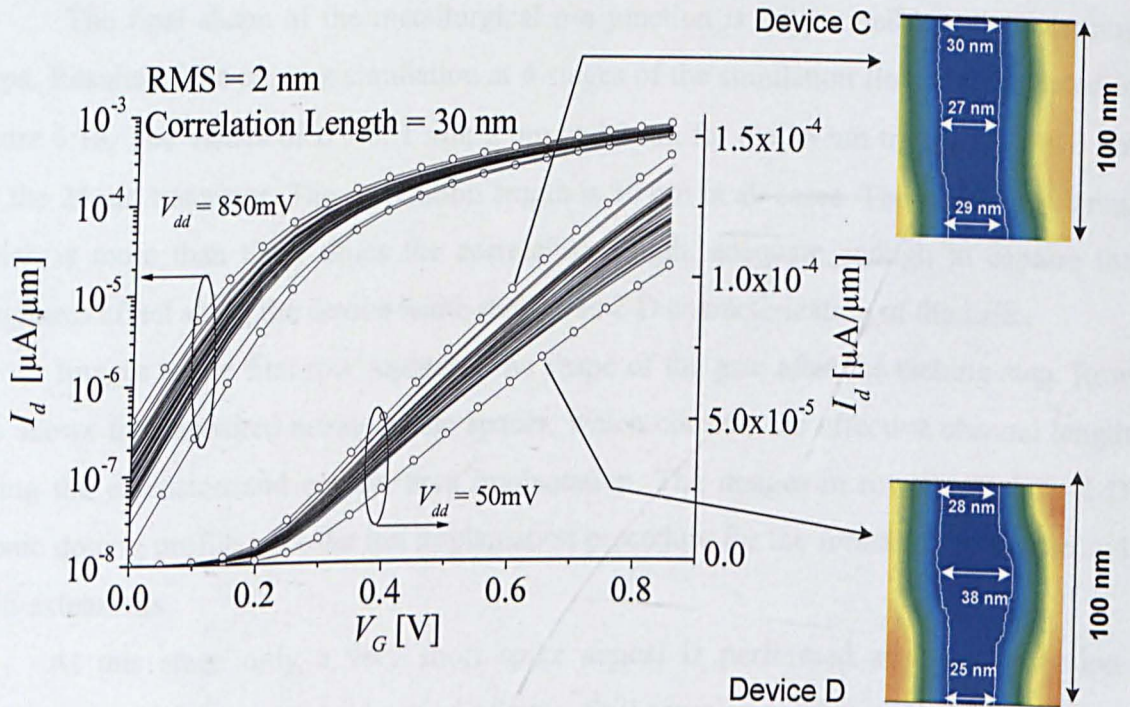


Figure 5:16 I_d - V_g characteristics of 50 n-channel MOSFET, RMS height of 2 nm.

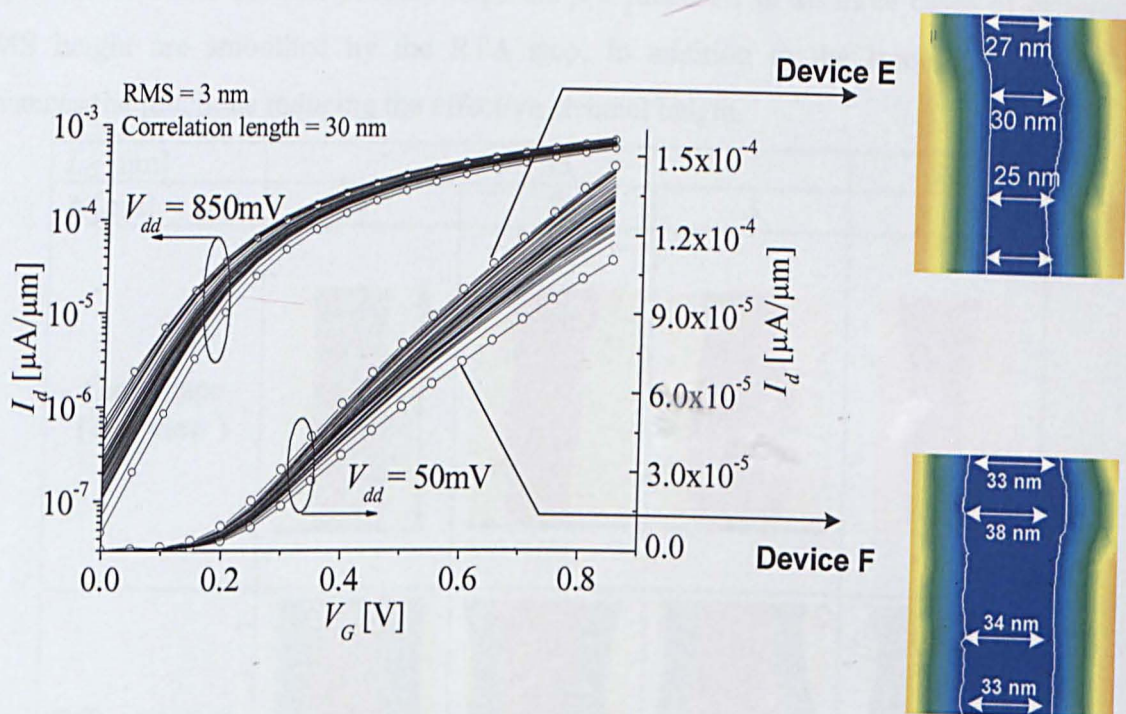


Figure 5:17 I_d - V_g characteristics of 50 n-channel MOSFETs, including LER effects with an RMS height of 3 nm. The image to the right depicts the two extreme cases occurring as a result of LER in devices E and F among the simulated 50 devices.

The final shape of the metallurgical p-n junction is determined by all processing steps. Results of the process simulation at 4 stages of the simulation flow are presented in figure 5:18. The values of Δ are: 1 nm, 2 nm and 3 nm for the 35 nm transistor, and 3 nm for the 25 nm transistor. The correlation length is 30 nm in all cases. The width is 100 nm, which is more than three times the correlation length, adequate enough to capture the roughness effect along the device width that gives 2-D characterization of the LER.

Images in the first row represent the shape of the gate after the etching step. Row two shows the deposited nitride offset spacer, which controls the effective channel length during the extension and contact area implantation. The images in row three show 2-D arsenic doping profile after the ion implantation procedure for the formation of source and drain extensions.

At this stage only a very short spike anneal is performed at a relatively low temperature and for a very short period of time. Still some smoothing of the metallurgical junction compared to the gate edge shape is observable. The strongest effect on the doping

distribution is the rapid thermal annealing (RTA) process after the source/drain contact implantation. After the full process steps the p - n junctions in all three cases of different RMS height are smoothed by the RTA step. In addition to the smoothing the RTA advances the junctions reducing the effective channel length.

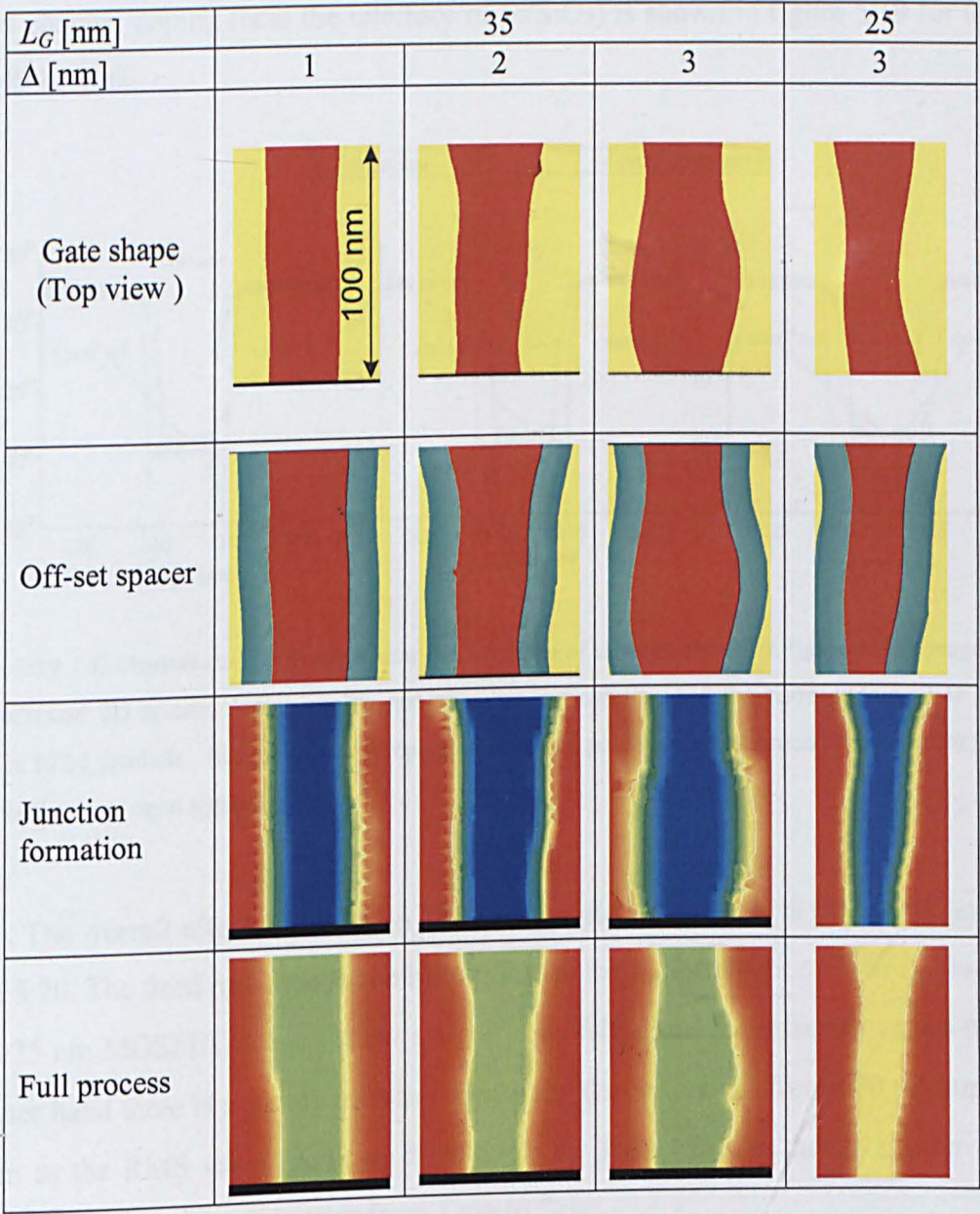


Figure 5:18 LER with different values of Δ . The red colour in pictures of row 1 and 2 represent the etched polysilicon gate material and yellow represents silicon dioxide insulator. The green colour in row 2 represents the silicon oxy-nitride spacer. The red colour in pictures of row 3 (after the formation of p - n junction where the polysilicon gate is removed for visualisation purpose) represent the arsenic (As) doping concentration in source and drain region and the blue colour represent minimal As concentration in the channel near the interface. The lines in pictures of row 3 and 4 show the position of p - n junctions. The width of the device is 100 nm

Figure 5:19 depicts samples of 1-D net doping profiles of the devices presented in figure 5:18. The net doping profile shown in closed and open symbols corresponds to devices in row three and four respectively. After full device process steps the increase in the net channel doping (near the interface of Si/SiO₂) is shown in figure 5:19 for all cases of RMS heights.

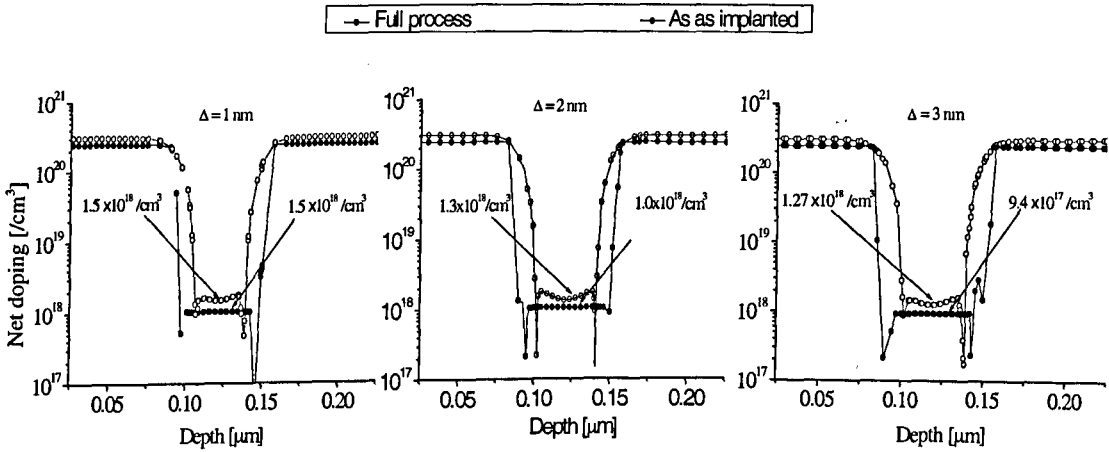


Figure 5:19 1-D channel doping profiles near the interface of devices shown in figure 5:18. Images in row three represent 2D arsenic doping profile just after the implantation and the corresponding 1-D profile is shown in filled symbols. Images in row 4 represent after full process simulation and the corresponding 1-D profile is shown in open symbols.

The overall effects of LER on IPF using various values of RMS are illustrated in figure 5:20. The third row depicts the histogram for the I_{on} distributions. The I_{on} fluctuation in the 35 nm MOSFET doesn't show a significant difference for different values of Δ . On the other hand there is a steady increase in average derive current from 670 $\mu\text{A}/\mu\text{m}$ to 741 $\mu\text{A}/\mu\text{m}$ as the RMS value increases from 1 nm to 3 nm. This is due to shorter channel length as the RMS value increases from 1 nm to 3 nm.

The corresponding off-current distribution is shown in row 1 and 2 both in linear and logarithm scales in figure 5:20. The off-current distribution often can be described as a lognormal. The fluctuation in off-current significantly increase with the increase of Δ . When $\Delta = 1, 2$ and 3 nm the standard deviation of the off-current is 55, 79, and 157 nA/ μm respectively (table 5:4). Since the roughness increases with increasing RMS values, the off current fluctuation is increasing as expected.

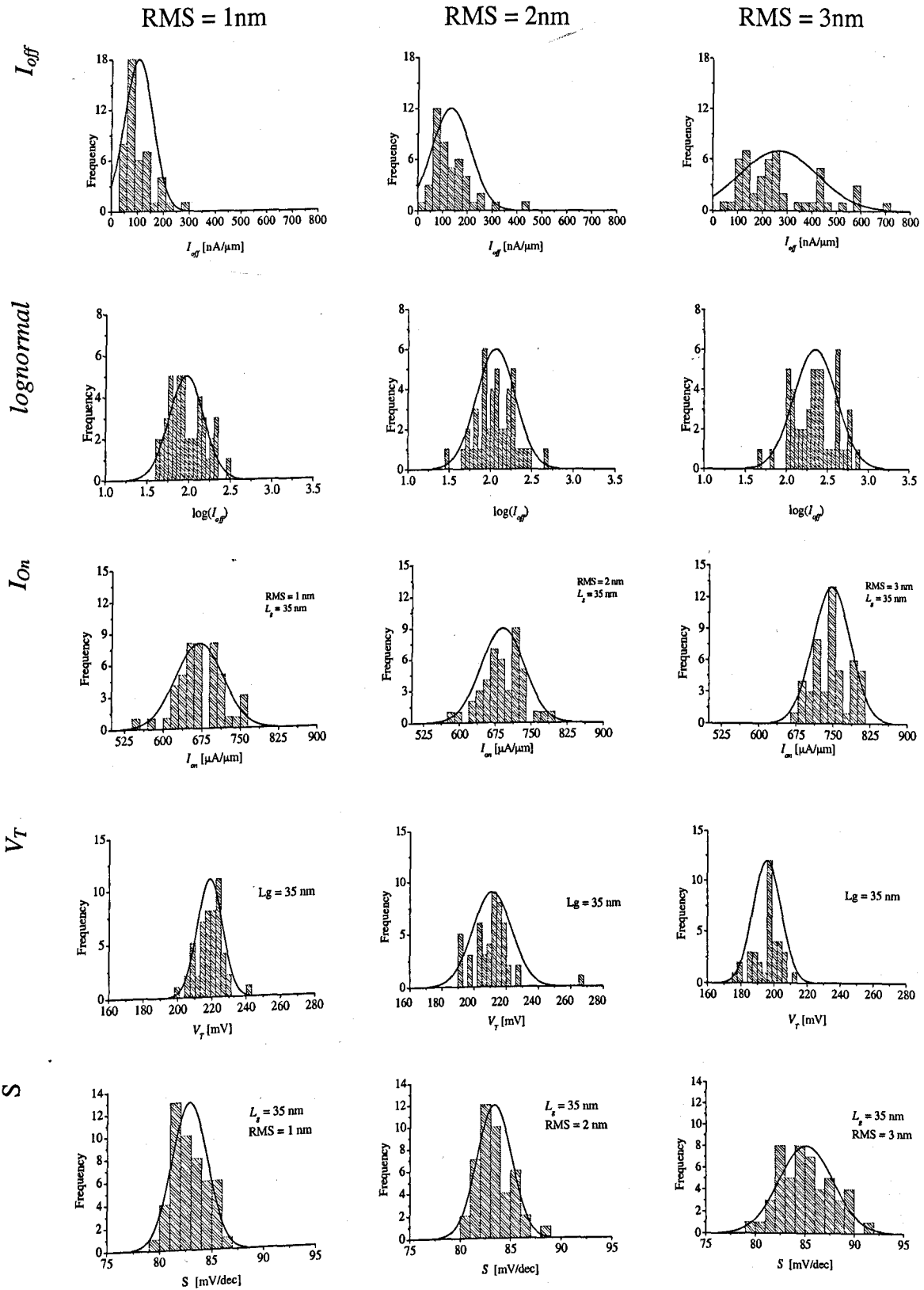


Figure 5:20 Histograms showing distributions of intrinsic device parameters in the 35nm n-MOSFET. The columns and rows represent RMS values and the corresponding distribution of electrical parameters

respectively. The correlation length is 30nm in all cases. Row 1 and 2 show linear scale and logarithmic scale of the *off*-current distributions respectively.

The negative impact of LER on the threshold voltage variations and subthreshold slope is also shown on the histogram distributions of V_T and S illustrated in figure 5:20. The two important features of electrostatic integrity (threshold voltage and subthreshold slope) in the 35 nm transistor are compared in row four and five. For example, the $\pm 3\sigma V_T$ variation for corresponding RMS values of 1, 2 and 3nm are $\pm(22, 30$ and $80\text{mV})$ respectively. The other aspect of V_T degradation is the short channel effect, which is the decrease of threshold voltage as the effective channel length is reduced [5.7]. As previously discussed, the length between the metallurgical p - n junction decreases for the The simulated values of important MOSFET parameters for the gate length of 35 nm and 25 nm with the effect of LER included are summarised Tables 5:3 and 5:4.

L_G	[nm]	35			25
Δ	[nm]	1	2	3	3
I_{on-MAX}	[$\mu\text{A}/\mu\text{m}$]	761	797	817	803
I_{on-MIN}	[$\mu\text{A}/\mu\text{m}$]	574	575	627	469
$I_{off-MAX}$	[nA/ μm]	285	411	710	2456
$I_{off-MIN}$	[nA/ μm]	41	14	46	334
I_{d-max} @ 50mV	[$\mu\text{A}/\mu\text{m}$]	137	143	146	99
I_{d-min} @ 50mV	[$\mu\text{A}/\mu\text{m}$]	97	104	110	158

Table 5:3 Comparison of electrical parameters 35 and 25nm gate length of n-channel MOSFETs with LER effect.

L_G	RMS	$\sigma(I_{off})$	$\sigma(I_{on})$	$\sigma(V_T)$	σS	$\langle I_{on} \rangle$	$\langle I_{off} \rangle$
[nm]	[nm]	[nA/ μm]	[$\mu\text{A}/\mu\text{m}$]	[mV]	[mV/dec]	$\mu\text{A}/\mu\text{m}$	nA/ μm
35	1	55	46	7.4	1.7	670	104
	2	79	45	10.03	1.5	690	131
	3	157	41	26.5	2.7	741	258
25	3	393	95	49.5	10.8	640	168

Table 5:4 Summaries of device parameters with the effect of LER

The $\pm 3\sigma I_{on}$ for the 25 nm MOSFET is about $\pm 285 \mu A/\mu m$ (see table 5:4) for the RMS value of 3 nm. Furthermore, comparing the 35 and 25 nm gate lengths MOSFETs, $\pm 3\sigma I_{off}$ variation in 25 nm transistor is approximately $\pm 1100 nA/\mu m$ which reflects a 60% higher off-current variation than in the 35 nm MOSFET.

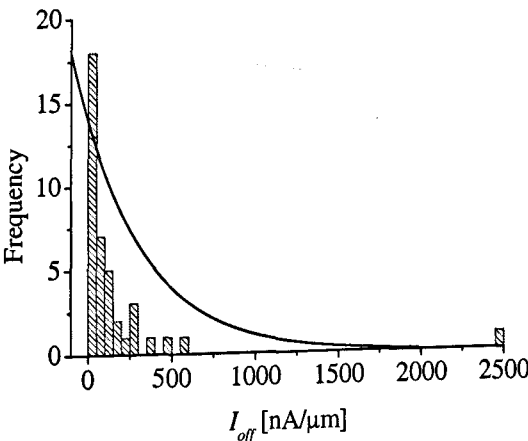


Figure 5:21 Linear scale of I_{off} distribution in the 25 nm transistor with RMS=3nm

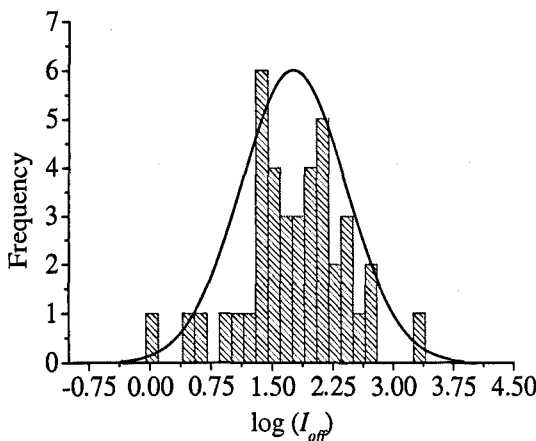


Figure 5:22 Logarithmic scale of I_{off} distribution in the 25nm transistor with RMS=3 nm.

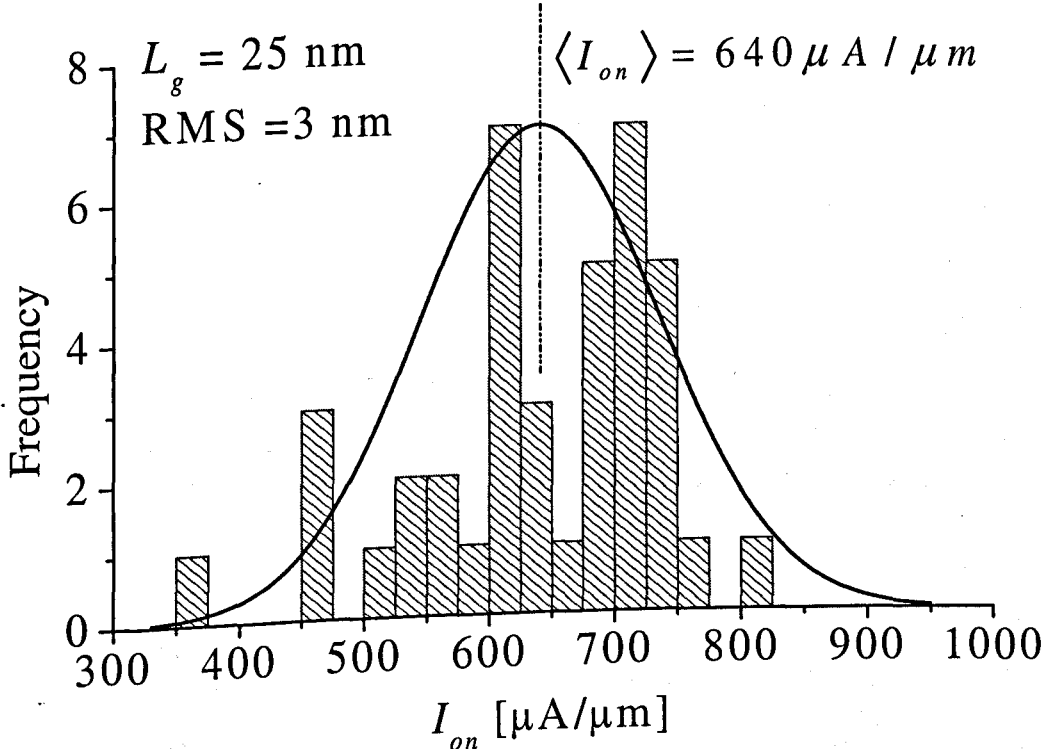


Figure 5:23 Histogram of drive current distribution for 25 nm transistor with RMS =3 nm.

5.2.4.2 The effect of correlation length (Λ)

As discussed earlier, one of the parameters used to characterise LER is the correlation length. In this section the effect of different values of Λ (5 nm, 15 nm and 30 nm) on metallurgical and p - n junctions and the device characteristics are presented. In all simulations here, the RMS value is kept 3 nm in order to separate the contribution of correlation length in characterizing the LER. The simulated devices have a 35 nm gate length. With the exception of different correlation lengths, the process corresponds to the calibrated 35 nm MOSFET.

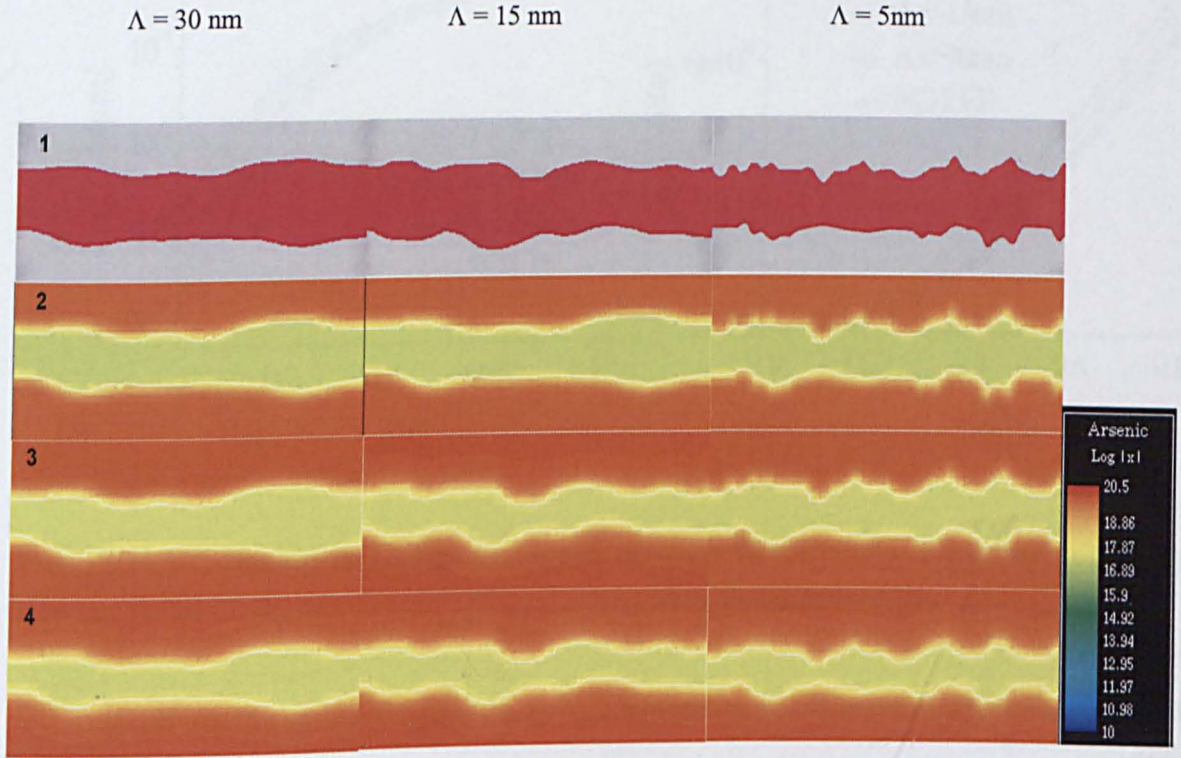


Figure 5:24 Gate shape and arsenic doping profile of three randomly generated MOSFETs using correlation lengths 30, 15, and 5 nm. Row 1 shows top view of gate shape before As-implantation (the grey colour represents nitride spacer and red represents polysilicon gate material) and rows 2-3 depict the mapping of arsenic concentration (orange colour), 5 seconds and then after 15 seconds annealing respectively and row 4 shows distribution of arsenic and p - n junctions after full process.

The gate shapes of three randomly generated MOSFETs using Λ of 5, 15 and 30 nm are illustrated at the top of figure 5:24 (first row). The arsenic doping profile and the metallurgical p - n junction immediately after the arsenic implantation, and at 5 seconds and after 15 seconds activation, are illustrated in rows 2 and 3 respectively.

The p - n junction position follows almost precisely the shape of the gate edge in the case of large correlation lengths. The 3D implantation and the diffusion associated with the activation step smear almost completely short wavelength features at small correlation lengths.

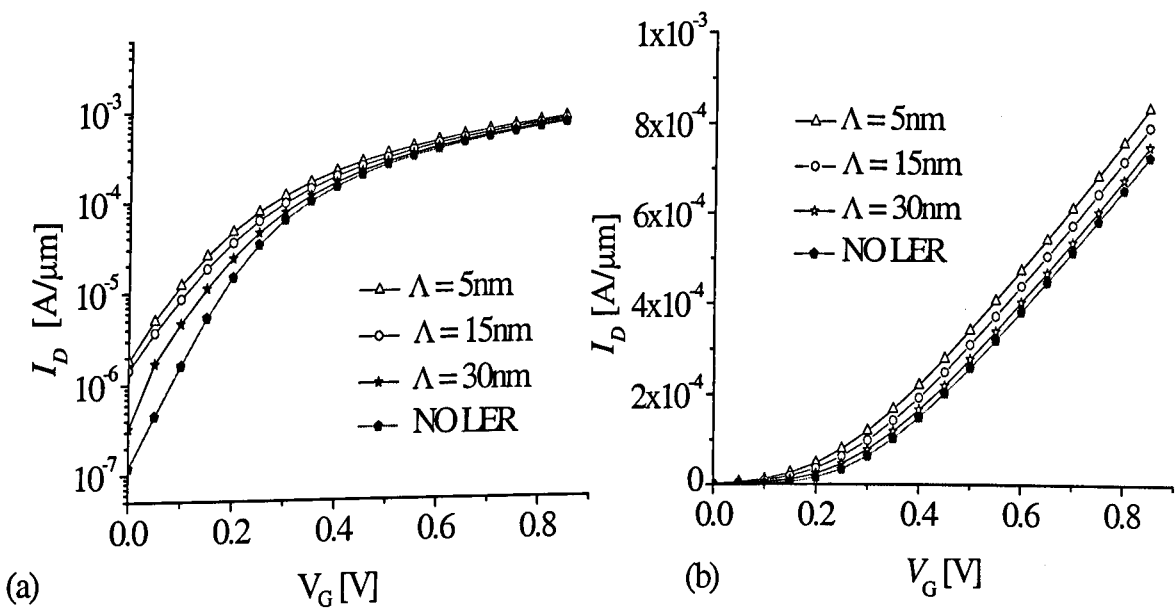


Figure 5:25 I_d - V_g Characteristics at high drain voltage of 850mV for the 35nm n -channel MOSFET, (a) log scale and (b) linear scale, with the effect of LER simulated for various correlation lengths

Figure 5:25a and b compare the I_d - V_g characteristics among three devices with LER and the reference 35 nm transistor without LER. Exactly the same doping conditions and process flow simulation have been applied to all four devices. The degradation of subthreshold slope is noticeable in figure 5:25a while the drive current variation can be deduced from figure 5:25b. When the correlation length is 5 nm I_{off} is almost an order of magnitude larger than the calibrated reference 35 nm device without LER.

Although I_{on} increases with decreasing correlation length, the more rapid increase in I_{off} results in an overall deterioration of device performance. The dependence of I_{off} and

I_{on} on Λ are summarized in table 5:5. The simulation results presented in this sub-section do not represent statistical measures of ensembles of devices (like the mean value and the standard deviation). Curves shown in figures 5:25a and 5:25b represent four individual devices. However, the results and trends of variations reasonably reflect the impact of different values of correlation length (Λ) on intrinsic device parameters.

Λ [nm]	I_{on} [$\mu\text{A}/\mu\text{m}$]	I_{off} [nA/ μm]	$\frac{I_{on}}{I_{off}}$	ΔI_{on} [$\mu\text{A}/\mu\text{m}$]	ΔI_{off} [nA/ μm]
10	836	1830	456	110	1730
20	794	1450	547	68	1350
30	751	330	2275	25	230
No LER	726	100	7255		

Table 5:5 Off-current and on-current in 35 nm n-MOSFET with LER of different correlation lengths

5.3 “Deterministic” LER simulation

In order to understand better the effect of LER/LWR on transistor performance and IPF we have simulated different specific patterns reflecting generic roughness configurations, which are referred in this work as “deterministic” LER. Although these specific geometric gate patterns do not represent the overall property of randomly generated rough edges, they can assist in investigating the effect of asymmetric gate edges on the I - V characteristics and the local distribution of dopants near the p-n junctions. It is also possible to determine which of the two irregular edges, source or drain, has the larger impact on device properties. In this section, the investigations of eight different combinations of geometric gate patterns are presented.

5.3.1 Local doping variation

The variation of gate length along the width of the device will affect the local doping distribution, in particular the pattern of the p - n junctions. Ideally the doping concentration will have the same profile everywhere along the width of the channel. This will result in a uniform effective channel length along the channel width. The uniformity results in uniform V_T , I_{on} and I_{off} . However, the presence of LER or LWR changes this idealistic picture of devices.

To highlight the impact of the gate pattern on the doping distribution, we examine the doping profiles of different positions along the channel width. One of the 3D device structures with a particular gate obtained from the process simulation was partitioned into four parts as illustrated in figure 5:26. The positions are selected where the doping profile along the channel was changed significantly as a result of the gate shapes changes. A vertical slice in those positions gives a corresponding 2-D image on x - y plane, from which the 1-D doping profiles near the interface are extracted.

Images of the 2-D doping profiles for arsenic in source and drain regions and the net doping profile in the device for all four different positions are shown in column one and two of figure 5:26. The cut plane positions are marked on the image that shows the top view of the gate. The profiles for both arsenic and net doping are not symmetric with respect to the middle of the channel.

For example, the image for the 2-D profile in row one of figure 5:26 depicts *fewer* dopants near the source junction than the drain junction which indicates that the doping distribution depends strongly on the 2D shape of the gate edge. The roughness of the gate edge, shown in figure 5:26b is passed onto the p - n junction during the implantation process but also smoothed by the high temperature annealing steps.

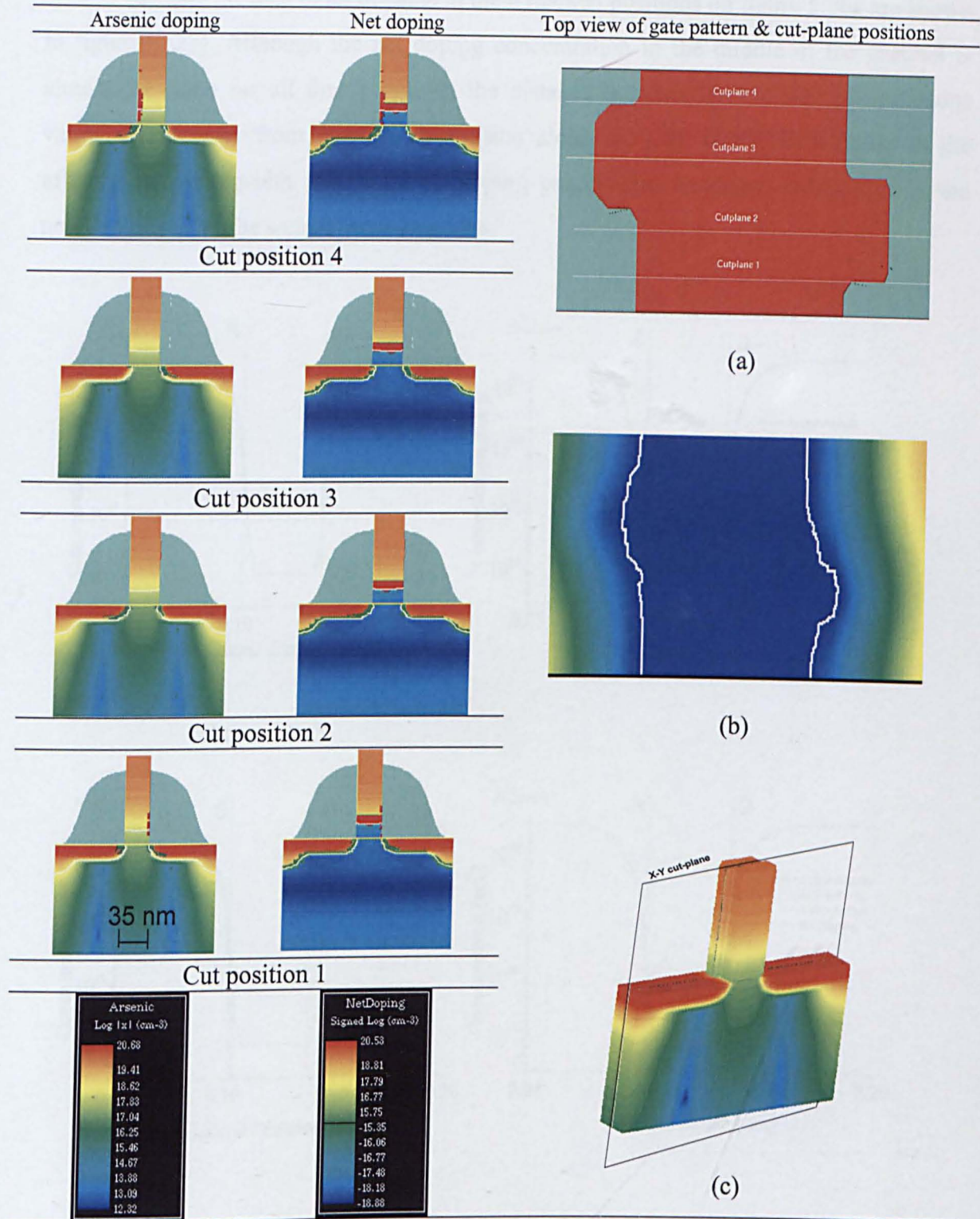


Figure 5:26 2-D doping profiles of arsenic dopants (first column, which is represented by red colour) and net doping profiles (second column where the blue colour represents all p -type and red represents all n -type dopants) in different positions along the device width. The figures in rows correspond to the doping profiles in the marked cut planes position (a), p - n junction (b), and the 3-D doping profile of arsenic is shown in (c). The net doping is usually defined as the difference between the acceptors (N_a , like boron and indium) and donors (N_d , like arsenic) dopants in the device.

The 1-D profiles of all dopants in the 4 marked positions on figure 5:26a are shown in figure (5:27). Although the net doping concentration in the middle of the channel is almost the same for all four positions, the distance between the metallic p - n junctions varies significantly from position to position along the gate width. In addition to the effective channel width variation, the doping profile also becomes asymmetric in the neighbourhood of the source/drain junctions.

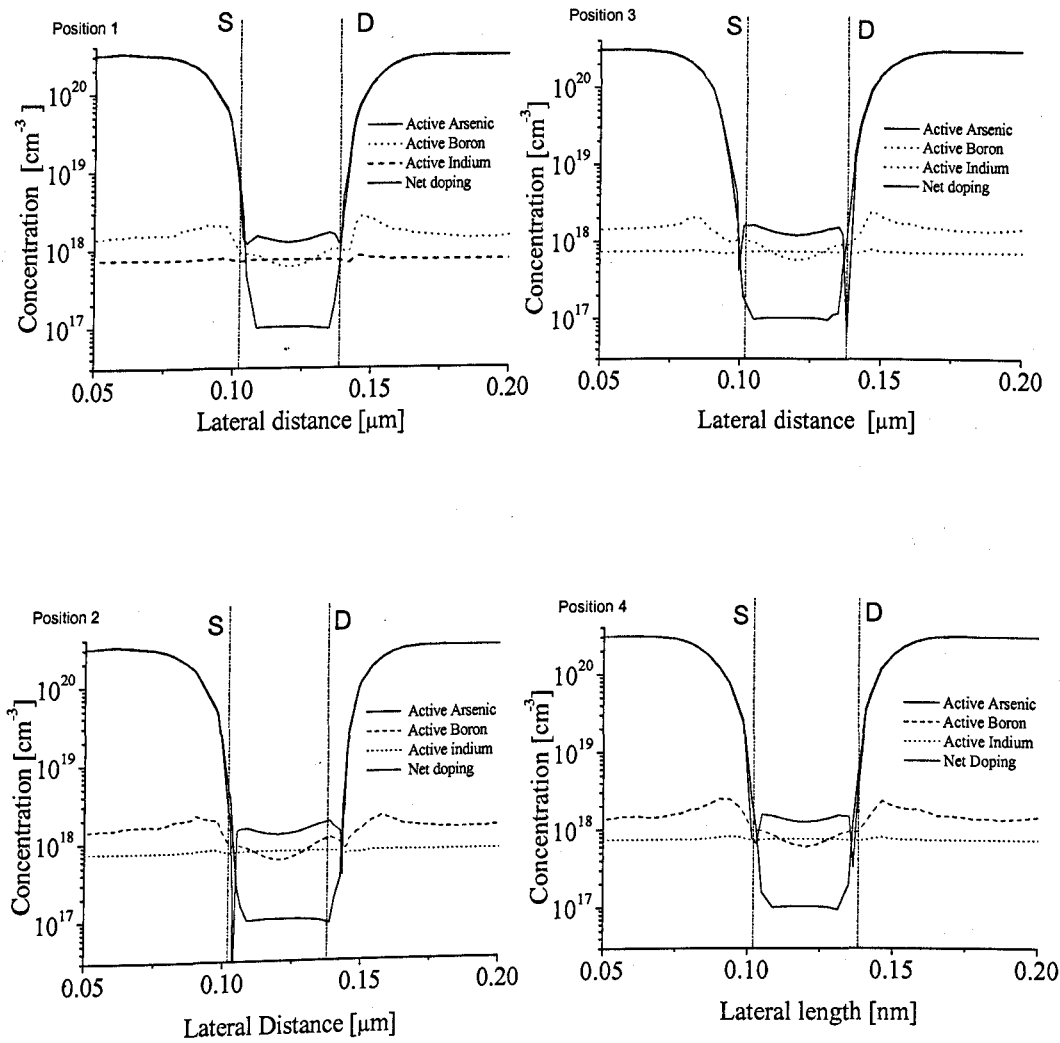


Figure 5:27 1-D doping profiles in the Si/SiO₂ interface along the channel of the same device but at different position along the gate width as shown in figure 5:26a.

5.3.2 The effect Deterministic LER on I-V characteristics

The eight specific geometric gate patterns which are employed to study the deterministic LER are shown in figure 5:28. The first row depicts the top view of the gate and the second row depicts the p-n junction of the corresponding geometric patterns after full process simulations.

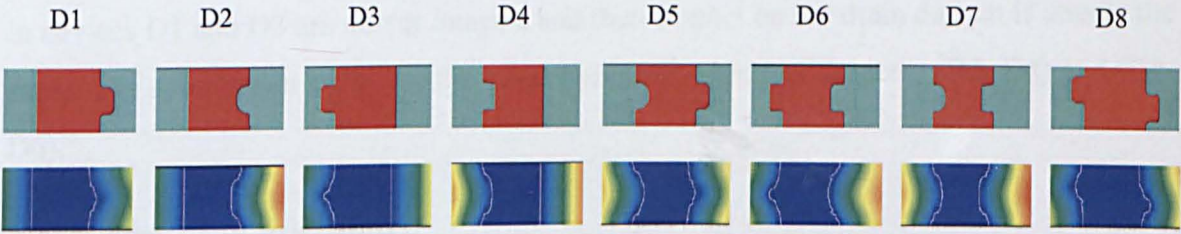


Figure 5:28 Basic geometric elements of LER

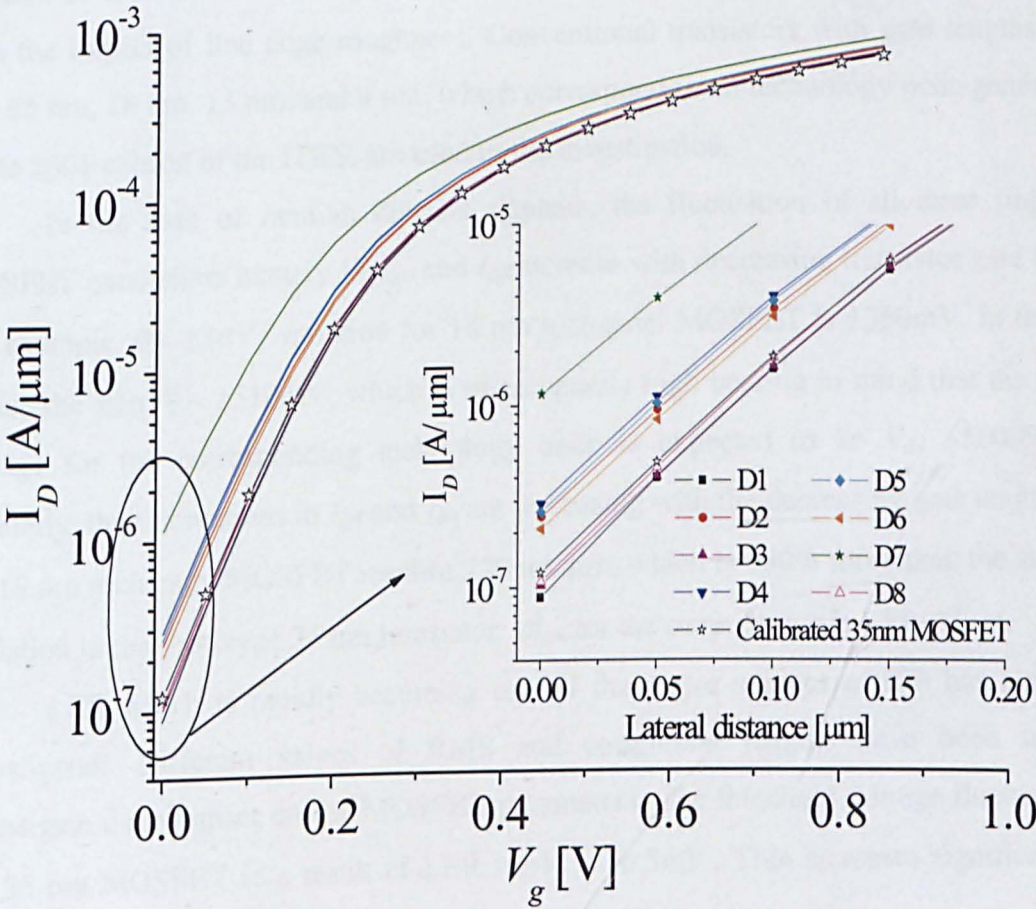


Figure 5:29 Id-Vg characteristics of the basic geometric gate patterns shown in figure 5:23

Device D7 has the worst $I_{off} = 1160 \text{ nA}/\mu\text{m}$, which is more than eleven times the off-current of the original prototype 35 nm MOSFET ($100 \text{ nA}/\mu\text{m}$). This is because the channel length in the middle of the device is greatly reduced, which results in an overall reduction of the effective channel length and as a consequence there is an increase in current. Another interesting observation is that it does not make a significant difference whether the defect is at the source or at the drain edge of the gate. For example the defects in devices D1 and D3 are mirror images, and their impact on the drain current is almost the same. The same result holds for the other two pairs of mirror devices, (D2, D4) and (D5, D6).

5.4 Chapter summary

Intrinsic parameter fluctuations in decanano MOSFETs induced by the number and position of discrete random dopants in the channel are statistically investigated, together with the impact of line edge roughness. Conventional transistors with gate lengths of 35 nm, 25 nm, 18 nm, 13 nm, and 9 nm, which correspond to all technology node generations in the 2001 edition of the ITRS, are used in the investigation.

In the case of random discrete dopants, the fluctuation of all three important MOSFET parameters namely V_T , I_{on} and I_{off} increase with decreasing transistor gate length. For example, the $\pm 3\sigma V_T$ variation for 18 nm n-channel MOSFET is $\pm 360 \text{ mV}$. In the 9 nm device the $\pm 3\sigma V_T \sim \pm 510 \text{ mV}$, which is unacceptably high bearing in mind that the supply voltage for the corresponding technology node is expected to be $V_{dd} \sim 500\text{-}600 \text{ mV}$. Similarly, the fluctuations in I_{off} and I_{on} are increasing with the decreasing gate length. σI_{off} for 18 nm n-channel MOSFET reaches $220 \text{ nA}/\mu\text{m}$, which is 180% more than the standard deviation in the prototype 35 nm transistor. σI_{on} for the same device is $\sim 45 \mu\text{A}/\mu\text{m}$.

LER which is rapidly becoming one of the major sources of IPF has also been investigated. Different values of RMS and correlation lengths have been used to investigate their impact on the MOSFET parameters. The threshold voltage fluctuation in the 35 nm MOSFET as a result of LER is $\sigma V_T = 26.5 \text{ mV}$. This increases significantly to $\sigma V_T = 45.5 \text{ mV}$, in the 25 nm MOSFET, an increase of 72%. The corresponding σI_{off} for the same devices are $157 \text{ nA}/\mu\text{m}$ and $393 \text{ nA}/\mu\text{m}$ respectively. However, in the 35 nm or 25 nm MOSFETs, random discrete dopants remain the dominant source of IPF.

5.5 Chapter references

- [5.1] A. Asenov, A. R. Brown, J. H. Davies, S. Kaya, and G. Slavcheva, "Simulation of Intrinsic Parameter Fluctuations in Decananometer and Nanometer-Scale MOSFETs", *IEEE Trans. Electron Device.* Vol. 50 (9), pp. 1837-1852, 2003
- [5.2] A. Asenov, A.R. Brown, J.H. Davies and S. Saini, "Hierarchical approach to 'atomistic' 3D MOSFET simulation", *IEEE Trans. on Computer Aided Design of Integrated Circuits an Systems*, Vol. 18, pp. 1558-1565 (1999).
- [5.3] Synopsys, Taurus process manual, Version V-2003.12, December 2003
- [5.4] S. Inab, K. Okano, S. Matsuda, M. Fujiwara, A. Hokazono, K. Adachi, K. Ohuchi, H. Suto, H. Fukui, T. Shimizu, S. Mori, H. Oguma, A. Murakoshi, T. Itani, T. Iinuma, T. Kudo, H. Shibata, S. Taniguchi, M. Takayanagi, A. Azuma, H. Oyamatsu, K. Suguro, Y. Katsumata Y. Toyoshima, and H. Ishiuchi, "High Performance 35 nm Gate Length CMOS With NO Oxynitride Gate Dielectric and Ni Salicide", *IEEE Trans. Electron Devices*, Vol. 49(12), pp. 2263 – 2270, 2002
- [5.5] A. Asenov, M. Jaraiz, S. Roy, G. Roy, F. Adamu-Lema, A. R. Brown, V. Moroz and R. Gafiteanu, "Integrated Atomistic Process and Device Simulation of Decananometre MOSFETs", *Proceedings of SISPAD2002*, pp., 95, 2002
- [5.6] R. F. Pierret, "Field Effect Devices", in Modular Series on Solid State Devices, Vol. IV, Edited by G. W. Neudeck and R. F. Pierret, 2nd Edition, Addison-Wesley, 1990
- [5.7] Y. Taur , and T. Ning, "Fundamentals of Modern VLSI Design", Cambridge University Press, Cambridge, UK, 1998
- [5.8] A. R. Brown, A. Asenov, and J. R. Watling, "Intrinsic Fluctuations in Sub 10-nm Double-Gate MOSFETs Introduced by Discreteness of Charge and Matter" *IEEE Transaction on Nanotechnology*, Vol. 1(4), December 2002
- [5.9] K. Hess, "Advanced Theory of Semiconductor Devices", Prentice Hall, New Jersey, USA 1988
- [5.10] S. M Sze, Semiconductor Devices, Physics and Technology, 2nd Edition, John Wiley, & Sons, Inc. 2002

-
- [5.11] K. Lakshmunar, R. A. Hadaway, and Miles A. Copeland, "Characterization and Modeling of Mismatch in MOS Transistors for Precision Analog Design", *IEEE J. Solid-State Circuit*, Vol. Sc.21(6), 1986
- [5.12] K. Takeuchi, R. Koh, and T. Mogami, "A study of the Threshold Voltage Variation for Ultra-Small Bulk and SOI CMOS", *IEEE Trans. Electron Devices*, Vol. 48(9), 2001
- [5.13] A. J. Bhavnagarwala, X. Tang, J. D. Meindl, "The impact of Intrinsic Device Fluctuations on CMOS SRAM Cell Stability", *IEEE J. Solid-State Circuit*, Vol. 36(4), pp. 658-665, 2001
- [5.14] B. Cheng, S. Roy, G. Roy, F. Adamu-Lema, A. Asenov, "Impact of Intrinsic Parameter Fluctuations in Decanano MOSFETs on yield and functionality of SRAM Cells", *Solid-State Electronics*, Vol.49, pp. 740-746, 2005
- [5.15] J.A. Croon, G. Storms, S. Winkelmeier, I. Pollentier, M. Ercken S. Decoutere, W. Sansen, and H.E. Maes, "Line Edge Roughness: Characterization, Modeling and Impact on Device Behavior", *IEDM Tech. Dig.*, pp. 307-310, December 2002
- [5.16] H. Namatsu, M. Nagase, T. Yamaguchi, K. Yamazaki, and K., "Influence of edge roughness in resist patterns on etched pattern", *J. Vac. Sci. Technol. B* 16(6), pp. 3315, 1998
- [5.17] J. Shin, Y. Ma, and F. Carrina, "Depth dependence of resist line-edge roughness: Relation to photoacid diffusion length", *J. Vac. Sci. Technol. B* 20(6), pp. 2927, 2002
- [5.18] J. S. Bendat, Principles and Application of Random Noise Theory, John Wiley and Sons, New York, 1958
- [5.19] S. M. Goodnick, R. G. Gann, J. R. Sites, D. K. Ferry, and C. W. Wilmsen, "Surface roughness scattering at the Si-SiO₂ interface", *J. Vac. Sci. Technol. B* 1(3), pp. 803-808, 1983
- [5.20] T. Yoshnoubu, A. Iwamoto, and, H. Iwasaki, "Scaling Analysis of SiO₂/Si Interface Roughness by Atomic Force Microscopy", *Jpn. J. Appl. Phys.* Vol. 33, pp. 383-387, 1994
- [5.21] J. S. Bendat, and Allan G. Piersol, Random Data: Analysis and measurement procedures, 3rd edition, John Wiley and Sons, New York, 2002

-
- [5.22] V. Constantoudis, G. Patsis, L. H. A. Leunissen, and E. Gogolides, "Line edge roughness and critical dimension variation: Fractal characterization and comparison using model function", *J. Vac. Sci. Technol. B* Vol. 22(4), pp. 1974, 2004
- [5.23] G. P. Patsis, V. Constantoudis, A. Tserepi, E. Gogolides, T. Hoffman, "Roughness Analysis of lithographically produced nanostructures: off-line measurement and scaling analysis", *Microelectronic Engineering* Vol. 67-67, pp. 319-325, 2003
- [5.24] V. Constantoudis, G. P. Patsis, A. Tserepi, and E. Gogolides, "Quantification of line-edge roughness. II. Scaling and fractal analysis and the best roughness descriptors", *J. Vac. Sci. Technol. B* Vol. 21(3), pp. 1019, 2003
- [5.25] J. A. Croon, L. H. A. Leunissen, M. Benndrof, R. Rooyackers, R. Rooyackers, K. Ronse, S. Decoutere, W. Sansen, and H. E. Maes, "Experimental investigation of the impact of line-edge roughness on MOSFET performance and yield", *Proceedings European Solid-State Device Research, 2003. ESSDERC '03*, pp. 227
- [5.26] A. Asenov, S. Kaya, and A. R. Brown, "Intrinsic Parameter Fluctuations in Decananometer MOSFETs introduced by Gate Edge Line Roughness", *IEEE Trans. Electron. Devices*, Vol. 50(5), pp. 1254-1260, 2003
- [5.27] J. Ruzyllo, *Semiconductor Glossary*, 1st edition, PROSTO Multimedia Publishing; 2004.

6

The scaling of silicon on insulator

In the previous chapters we discussed the scaling of conventional MOSFETs, and the limitations and challenges that the semiconductor industry is facing to sustain Moore's law until 2018 (end of the present roadmap) and to extend it beyond. It was highlighted that conventional transistors with gate length of 18, 13, and 9 nm are increasingly difficult to design due to severe short channel effects, gate oxide tunnelling, power dissipation, intrinsic parameter fluctuations and the resulting loss of performance.

Therefore, it is very important to consider alternative device architectures that do not have some of the limitations of conventional MOSFETs and that may be realised with minimal changes to process technology. That is why the last two editions of the ITRS cover emerging research devices and technologies in great detail. At the same time, it calls for the attention of the research communities and the microelectronics industry to intensify research and development on non-classical FETs.

According to the 2004 update of the ITRS [6.1], there are four other categories of non-classical CMOS structures aside from the 'quasi-ballistic' FETs [6.2][6.3]. The first category is transport enhanced FETs [6.4], which allow improvements in current drive by using mechanically strained inversion layers or by introducing materials into the channel with higher carrier mobility than silicon (for example the use of III-V compound semiconductor materials) [6.5].

The second type of architecture is ultra thin body SOI FETs [6.6]. The other two categories are: source/drain engineered FETs (to minimise parasitic resistance and capacitance in source and drain regions) [6.7], and multiple gate FETs [6.] with the potential advantages of high drive current and improved short channel effect and subthreshold slope.

Dealing with all the categories of emerging devices and technology mentioned above would require multiple research programmes. In this work we concentrated solely on the scaling of ultra thin body silicon on insulator (UTB SOI) FETs. The results have been used by other members' of Device Modelling Group at the University of Glasgow to study intrinsic parameter fluctuations in these scaled devices and their subsequent impact on SRAM cells.

This chapter is structured as follows. The first section provides a brief description of the generic structure, the basic properties, and the advantages of SOI MOSFETs. The second section discusses scaling scenarios for the UTB SOI MOSFETs considered in this work. The electrical characteristics corresponding to various scaling scenarios are presented in section three and concluding remarks are given in the final section.

6.1 SOI device structure

SOI transistors have been around for more than two decades [6.8]. However, poor performance and instability in the early stages of their development, together with the continued scalability of classical MOSFET structure, kept SOI transistors out of favour in the mass market of semiconductor industry.

The slow progress in improving SOI wafers was related to problems with uniformity of the thin silicon layer, defect density at the bottom interface and problematic bonding technology, among other technological difficulties. From a circuit design point of view, the floating body effect was a major problem in partially depleted SOI devices. All of this changed in the early 90's when SOI devices began to promise the capability of delivering better performance than the conventional transistors [6:9]. With the introduction of smart cut and Unibond technology [6:10][6:11], the possibility of manufacturing high volume 300mm SOI wafers become a reality [6:12].

The demonstration of encouraging research devices and, in particular, ultra-thin body (UTB) SOI MOSFETs [6.13] increased the chance of such devices replacing the conventional MOSFETs at the 45nm technology node and beyond.

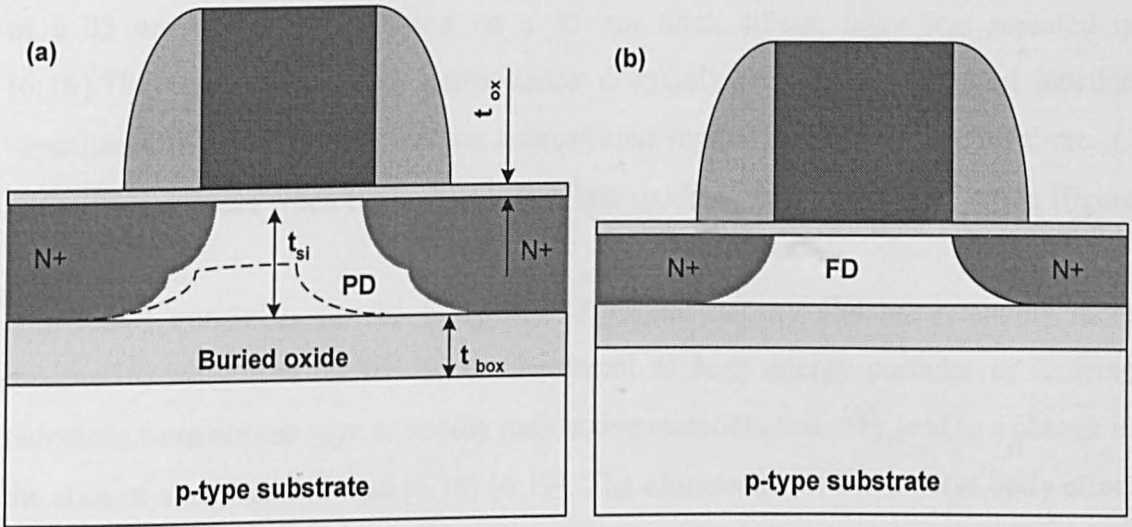


Figure 6:1 Typical PDSOI (a) and FDSOI (b) structures

The structure of the classical SOI transistor is similar to the structure of a conventional MOSFET with the exception of accommodating a buried oxide in the silicon substrate as shown in figure 6:1 [6.14]. The main reason why the oxide is placed just under the diffusion depth of the implanted impurities in the source/drain region is to eliminate (minimize) the junction capacitance. The reduction of junction capacitance increases the switching speed of the devices in CMOS circuits

6.1.1 Partially depleted and fully depleted SOI

There are two types of SOI transistors, partially depleted (PD) and fully depleted (FD), illustrated in figure 6:1a and 6:1b respectively. This section reviews the major properties and shortcomings of both PD and FD SOI MOSFETs. In PD SOI the silicon body is thicker than the gate depletion width so that the depletion layer covers partially the space under the gate, while in FD SOI the silicon layer thickness is smaller or equivalent to the depletion width; hence the name fully depleted SOI.

PDSOI has some advantages over conventional MOSFETs. For example, PDSOI circuits have higher operating speed. They can deliver up to 20% more performance than the corresponding bulk transistor and can reduce power dissipation for the same supply voltage by up to the 50% [6:15]. A 25% performance improvement of a 33 nm transistor fabricated on a 45 nm thick silicon layer was reported in [6:16]. The increase in circuit performance is mainly attributed to reduced junction capacitance (C_j) [6:17] between the source/drain region and the silicon substrate. C_j virtually disappears when the insulator (the box oxide) is buried in the substrate (figure 6:1)

SOI transistors can be integrated to greater density and are relatively more resistant to soft-errors caused by bombardment of high energy particles or ionizing radiations from cosmic rays or nearby radioactive materials that may lead to a change in the state of stored digital data [6:18] [6:19]. The elimination of the reverse body effect in stacked digital circuit applications [6:20] is also another advantage of SOI over bulk transistors.

However, PDSOI has its own disadvantages as a result of the way the device is structured. One of the major shortcomings of PDSOI is the floating body effect. It is the source of most of the unwanted properties of SOI transistors, for example: unstable threshold voltage (dynamic V_T variation) due to the variation of substrate-source voltage (V_{BS}) [6:9] [6:21], a sudden increase of the drain current near the saturation region, often termed as a “kink effect”[†] [6:22] [6:23] and the switching memory effect which may cause delay variations in circuits.

Among the suggested solutions for minimizing the floating body effect is a reduction in silicon thickness, t_{si} , so that there will not be a neutral region, or to use a body contact [6:24]. Creating a body contact (body-grounding) is unfavourable from manufacturing point of view. The alternative approach, the reduction in t_{si} leads to the second type of SOI structure known as fully depleted (FD) [6:25]. In FD SOI, the silicon layer is smaller than the gate depletion width. The absence of an un-depleted substrate greatly reduces the floating body effect in FD SOI.

[†] Kink effect is mainly caused by impact ionization resulting in the increase in V_{BS} with increasing V_{dd} .

6.1.2 Ultra thin body SOI

Ultra thin body (UTB) SOI is the extension of FDSOI. The main difference between the two is the silicon layer thickness. The silicon layer thickness for UTB-SOI transistors is usually, $t_{si} \leq 10nm$ [6.26] [6.27]. UTB SOI FETs are expected to deliver a solution to the difficult challenge facing the scaling of conventional MOSFETs to deep nano-scale dimensions [6.7]. In the past there was some concern about carrier mobility degradation for p-type UTB SOI for small t_{si} [6.28].

However recent work in [6.29] shows that UTB SOI transistors are a promising alternative to the conventional MOSFET near the end of the ITRS. Furthermore; recently researchers have experimentally demonstrated excellent carrier mobility characteristics in p-type UTB SOI by changing the silicon crystal lattice orientation from the conventional (100) to (110) [6.30].

The specifications for silicon layer thicknesses shown in table 6:1 are from the 2001(highlighted) and 2003 editions of the ITRS. They are basically identical, bearing in mind that the values given in the 2003 edition are for the “*initial*” thickness of the silicon layer, while in the 2001 edition, the dimensions are given for the “*final*” silicon layer thickness in the FD SOI wafer fabrication stages.

Year of production	2004	2007	2010	2013	2016
Technology Node	90	65	45	32	22
Physical L_g	37	25	18	13	9
ITRS’03 (Initial - t_{si})	21-39	18-33	15-19	14-16	13-14
ITRS’01 (Final - t_{si})	(11-19)	(8-13)	(5-9)	(4-7)	(3-5)
ITRS’03 (Initial t_{box})	56-94	42-70	26-44	18-32	14-22
ITRS’01(Final t_{box})	(28-46)	(19-31)	(14-23)	(10-16)	(7-11)

Table 6:1 ITRS requirements of the silicon layer and SiO₂ thickness in 2001 and 2003 editions (all dimensions in nm).

6.2 Scaling Study of UTB-SOI MOSFETs

The scaling study of the UTB SOI here is mainly focused on devices with channel lengths required for the 32 and 22 nm technology nodes (see table 6:1) and beyond. The generic device structure together with important design parameters used for device simulation is illustrated in figure 6:2. Initially, the scaling investigations of UTB-SOI transistors were performed on wide range of channel lengths spanning from 5 nm to 15 nm as shown in table 6:2, for various combination t_{si} and t_{ox} values.

The starting point in the scaling investigation is a 15 nm device with silicon body thickness of $t_{si} = 3$ nm and gate oxide of $t_{ox} = 1$ nm. The value of silicon layer thickness $t_{si} = 3$ nm has been selected based on the finding that electron mobility increases when t_{si} is in the range of 3-5 nm, due to sub-band level modulation by quantum mechanical confinements effects in ultra thin Si layers [6:31] [6:32] [6:33]. Table 6.2 shows the various scaling schemes where the values for t_{si} and t_{ox} are approximated based on equations (6.1) and (6.2)

$$t_{si}^* \approx \left(\frac{L_c^*}{L_{15}} \right) t_{si} \quad (6.1)$$

$$t_{ox}^* \approx \left(\frac{L_c^*}{L_{15}} \right) t_{ox} \quad (6.2)$$

L_c^* is the channel length of the scaled thin SOI MOSFETs. The investigation of the electrical parameters of the scaled UTB-SOI MOSFETs, presented in section 6.2.1 is performed using the device simulator, Medici [6:34]. Concentration dependant, lateral and vertical electric field dependant mobility models have been employed to perform the device simulation.

Once the electrical properties of devices shown in table 6:1 were simulated, three devices with channel lengths 15 nm, 10 nm and 5 nm were selected for further scaling investigations. The silicon layer thicknesses of these devices have been sub-linearly scaled assuming the outline given in table 6:2, the ratio between two successive channel lengths. The value of t_{si} for the corresponding channel lengths of 15 nm, 10 nm and 5 nm SOI are 3, 2.5 and 2 nm and the oxide thickness was scaled linearly. During

the scaling process, both t_{Si}^* and t_{ox}^* of the 15 nm channel device, were kept constant of 3 nm and 1 nm respectively as a reference device parameters. The highlighted values in the last row of table 6:2 depict the three selected devices.

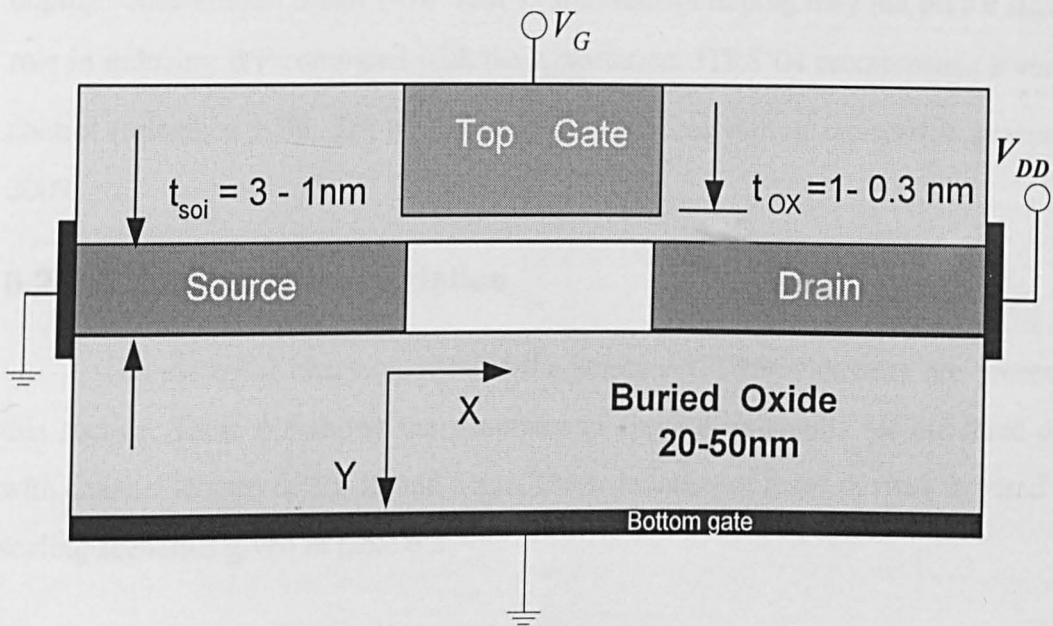


Figure 6:2 Generic diagram of ultra-thin SOI used for investigations of various scaling schemes in this work.

L_c [nm]	15		12.5		10		7.5		5	
Scaled parameters [nm]	t_{Si}	t_{ox}	t_{Si}	t_{ox}	t_{Si}	t_{ox}	t_{Si}	t_{ox}	t_{Si}	t_{ox}
Not scaled	3	1	3	1	3	1	3	1	3	1
t_{ox} : linearly scaled t_{Si} : constant	3	1	3	0.8	3	0.7	3	0.5	3	
t_{ox} : constant t_{Si} : linearly scaled	3	1	2.5	1	2	1	1.5	1	1	1
t_{ox} : linearly scaled t_{Si} : linearly scaled	3	3	2.5	0.8	2	0.7	1.5	0.5	1	0.3
t_{ox} : linearly scaled t_{Si} : Sub-linearly scaled	3	1			2.5	0.7			2	0.3

Table 6:2 Summaries of investigated scaling scenarios UTB-SOI MOSFET

The major scaling issue of UTB-SOI is control of the silicon layer thickness. The fluctuation of device parameters in UTB-SOI comes from interface roughness in both sides of the thin silicon layer [6.35]. Since the channel is un-doped - or at least the doping concentration is low ($\sim 10^{15}/\text{cm}^2$) - the random doping may not play a significant role in inducing IPF compared with the t_{si} variation. ITRS'04 recommends a very tight control (tolerance $\pm 5\%$, 3σ) of silicon layer thickness during fabrication processes by 2009.

6.2.1 Electrical characteristics

The electrical characteristics of the selected UTBSOI devices are presented in this section. Table 6:3 shows the summary of simulation results for the three devices with channel lengths of 15, 10 and 5 nm. The simulation of these devices is based on the scaling scenarios given in table 6.2.

L_c [nm]	t_{soi} [nm]	t_{ox} [nm]	V_T [mV]	ΔV_T [mV]	S [mV/dec]	I_{on} [$\mu\text{A}/\mu\text{m}$]	I_{off} [nA/ μm]	I_{on}/I_{off}
15	3	1	262	119	91	1283	88	14535
10	3.0	1.0	182	305	131	1541	25900	59
	3.0	0.7	223	103	113	1823	1547	1178
	2.0	1.0	241	167	99	1286	616	2088
	2.0	0.7	268	170	87	1552	70	22171
	2.5	0.7	246	158	98	1693	692	2447
5	3.0	1.0	-42	1100	231	2863	9×10^5	3
	3.0	0.3	104	507	204	3803	4×10^5	10
	1.0	1.0	168	1067	115	2871	1×10^6	3
	1.0	0.3	292	67	73	1861	3	6×10^5
	2.0	0.3	209	219	111	2955	9×10^3	347

Table 6:3 Investigations of various scaling scenarios

The first four rows of for 10 and 5 nm devices correspond to linear scaling described in the previous section. By taking in to account the balance between the performance factor (current drivability) and electrostatic integrity (good V_T , S , and I_{off}), the forth option shown for both devices is the optimum choice. The ratio I_{on}/I_{off} for

both devices is about four order of magnitude which is acceptable margin between the off and drive currents. The subthreshold slope of 5 nm device for the fourth option is near ideal (73mV/dec), and the subthreshold slope for the 10 nm device is also very good (87mV/dec).

However, manufacturing a silicon layer of 1 nm (roughly 2 lattice constants) for the 5 nm device is an extremely difficult technological challenge. Therefore we opted for the pragmatic scaling scheme highlighted in table 6:3 where the oxide thickness is scaled linearly along the gate length while the silicon layer thickness is scaled sub-linearly.

The I_d - V_g characteristics of the three devices at $V_d=50$ mV is plotted using a linear scale in figure 6.3. The shape of the characteristics clearly indicates increasing problem associated with the access resistance of the scaled devices. To highlight the subthreshold behaviour of the three devices, the semi-logarithmic plot of the I_d - V_g curves at low drain voltage of $V_d = 50$ mV is shown in figure 6.4. The subthreshold swing of 15 and 10 nm devices is reasonably good compared to the estimated $S = 111$ mV/dec for the 5 nm device *where*. The threshold voltage V_T for all three devices adjusted by the gate work function is about one quarters the supply voltage of $V_{DD} = 850$ mV, which is within acceptable device design rules.

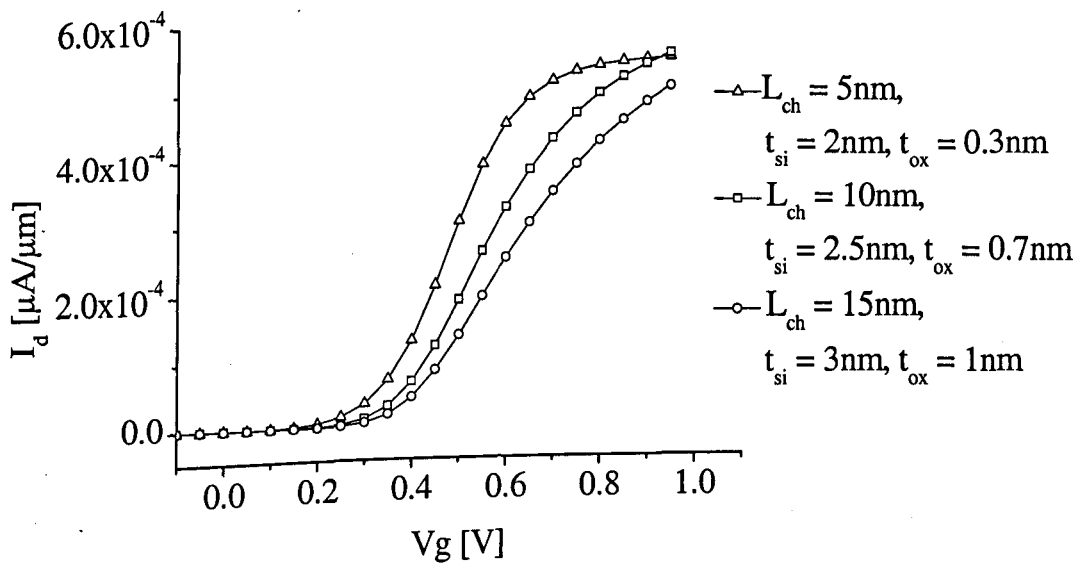


Figure 6:3 Plot of linear scale I_d - V_g characteristics at low drain voltage of $V_{DD} = 50$ mV of the same devices.

The high drain voltage I_d - V_g characteristics for all three devices are also plotted in figure 6:4. The 15nm channel UTB SOI transistor has a drive current of 1283 $\mu\text{A}/\mu\text{m}$ at the supply voltage of $V_{dd} = 850\text{mV}$ (figure 6:5a). In the case of the 10 and 5 nm channel SOI devices there is an increase of 32% and 130% in I_{on} compared with the 15 nm channel SOI device. However, the off-current increase in the 10 and 5 nm channel devices is much higher than in the case of the 15 nm channel device (Figure 6:5b). This can be adjusted by an appropriate choice of threshold voltage, at the expense of reduced current drive.

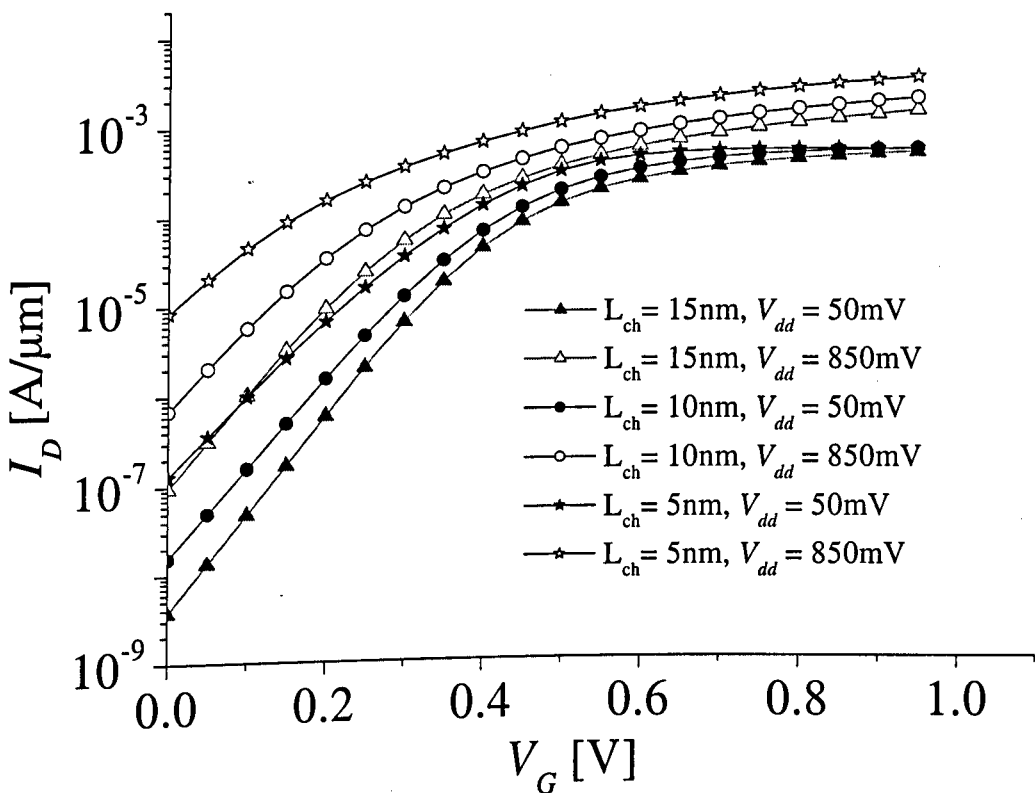


Figure 6:4 I_d - V_g characteristics of the UTB-SOI with $L_c = 15, 10$ and 5nm.

Figures 6:5 and 6:6 depict channel length dependence of the critical performance parameters, and electrostatic properties of the three devices respectively. As expected The 5 nm transistor has the worst parameters in almost all categories. Both the off-current and the threshold voltage roll-off, ΔV_T , are high for the 5 nm device compared with the other two devices.

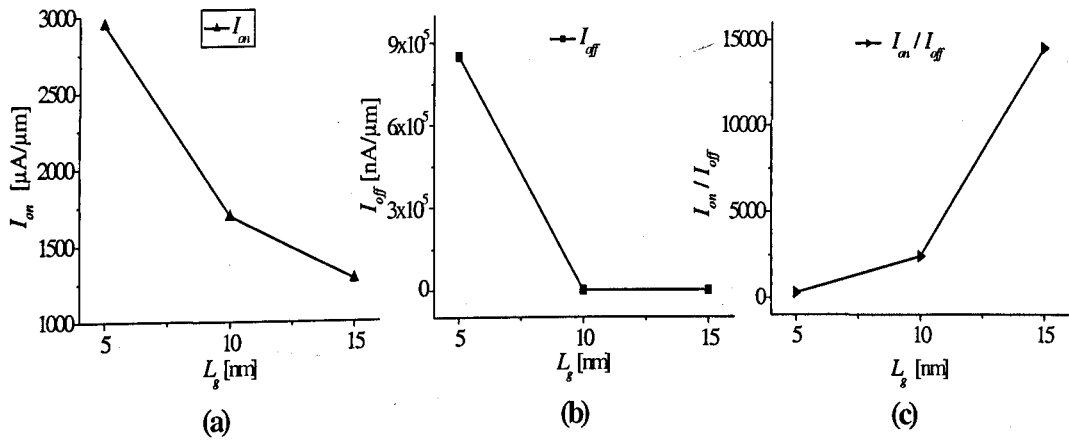


Figure 6:5 Critical performance parameters

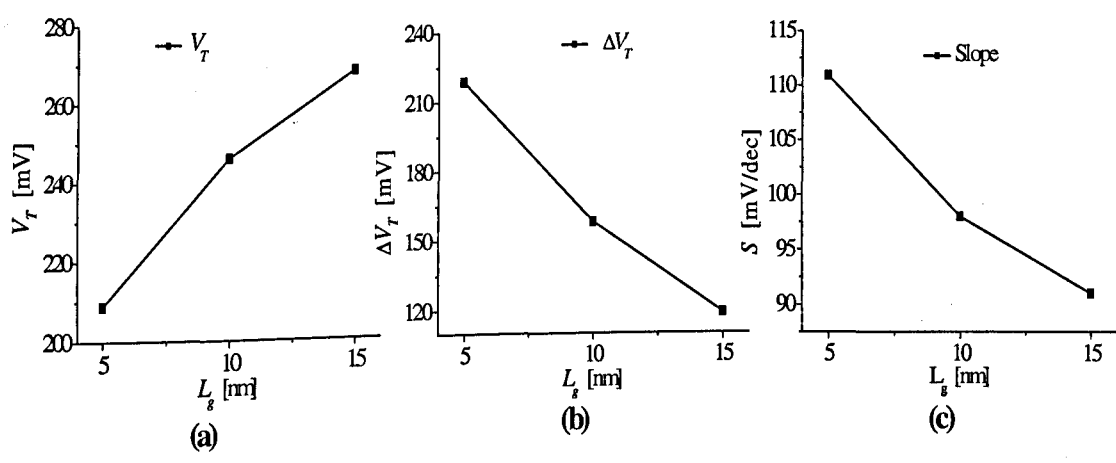


Figure 6:6 V_T , ΔV_T and subthreshold slope

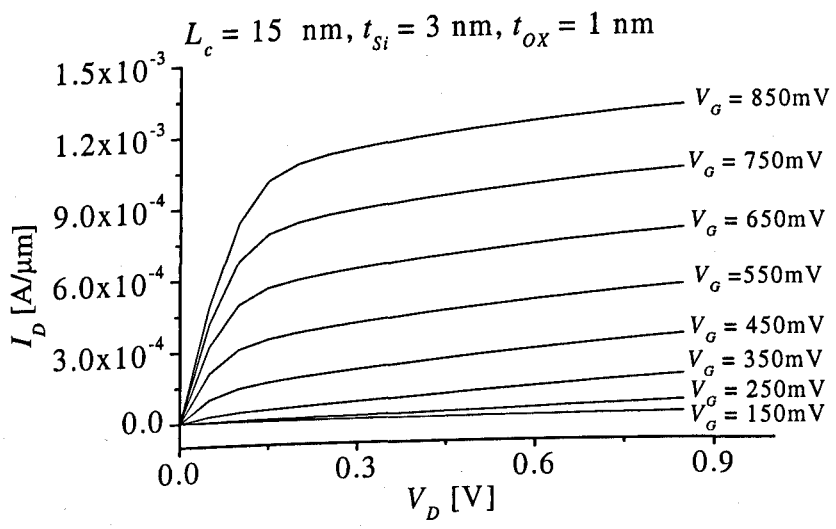


Figure 6:7 I_d - V_d (a) and I_d - V_g (b) characteristics of the 15 nm channel

For completeness I_d - V_d characteristics for the three devices with 15 nm, 10 nm, and 5 nm channel lengths are given in figures 6:7-6:9. In all three cases the output characteristics show well defined current saturation.

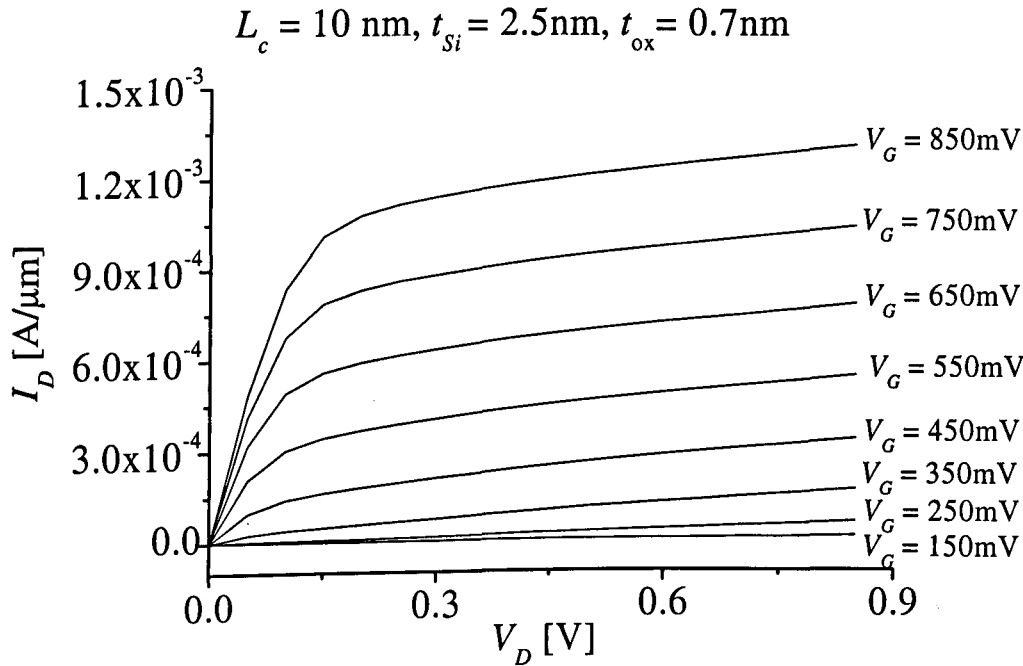


Figure 6:8 I_d - V_d (a) and I_d - V_g (b) characteristics of the 10 nm channel

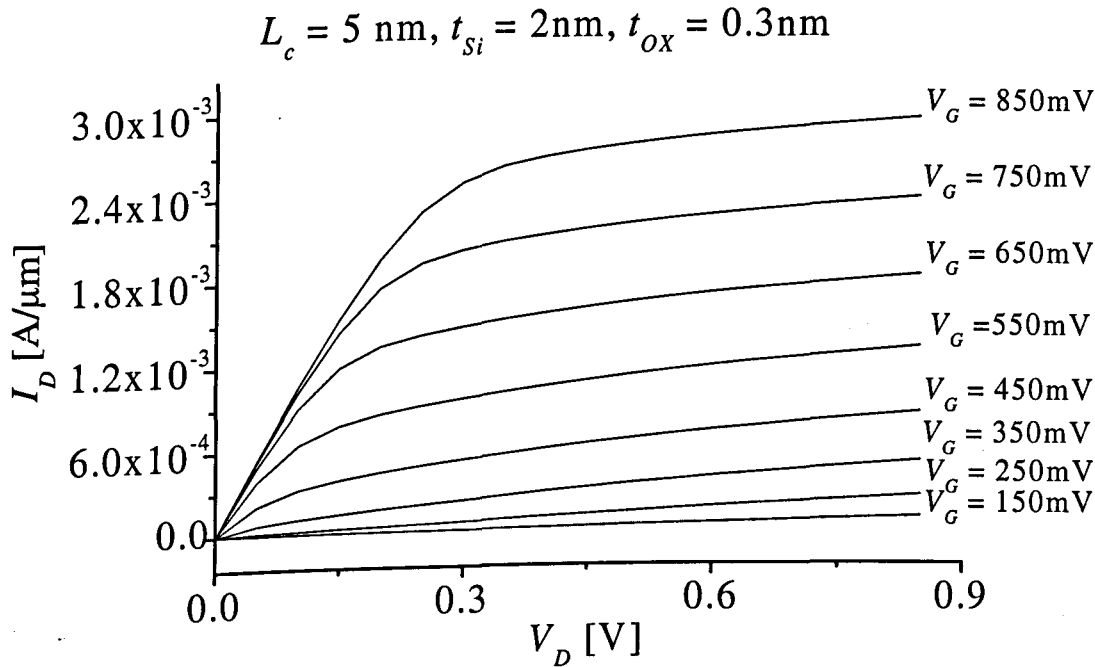


Figure 6:9 I_d - V_d (a) and I_d - V_g (b) characteristics of the 5 nm channel

6.2.2 Mobility and carrier velocity in UTB SOI

Electron mobility, electric field and the electron velocity along the channel of the selected 3 devices with 15 nm, 10 nm, and 5 nm are depicted in figure 6:10 a, b and c. As expected from drift-diffusion simulations, the maximum electron velocity in the channel of these devices is around the saturation electron velocity in the bulk silicon ($\sim 10^7\text{cm}^2/\text{V}\cdot\text{s}$). This is an indication that the simulation, which does not include velocity overshoot effects, underestimates device performance.

Further reasons to expect higher performance in the simulated devices is the absence of ‘volume inversion’ [6:36] and confinement related lifting of the degeneracy of the 6 fold Δ -valley in silicon.

The doping concentration in the channel ($10^{15}/\text{cm}^3$) is very low compared to the equivalent conventional MOSFET ($\sim 10^{19}/\text{cm}^3$) discussed in chapter 4, and results in a significant increase in the mobility and the carrier velocity in the simulated devices. The approximate average electron mobility along the channel of all three is plotted on figure 6:10a. Comparing the carrier mobility of, for example, the 10 nm UTB SOI device with a scaled conventional 9 nm MOSFET (chapter 4); there is more than a 100% mobility enhancement (see table 6:4).

	L_g [nm]	$\langle\mu_e\rangle$ [$\text{cm}^2\text{V}^{-1}\text{s}^{-1}$]
SOI MOSFET	5	59
	10	74
	15	78
Conventional MOSFET	9	21
	13	32
	18	35

Table 6:4 Carrier mobility in UTB SOI and scaled conventional MOSFETs

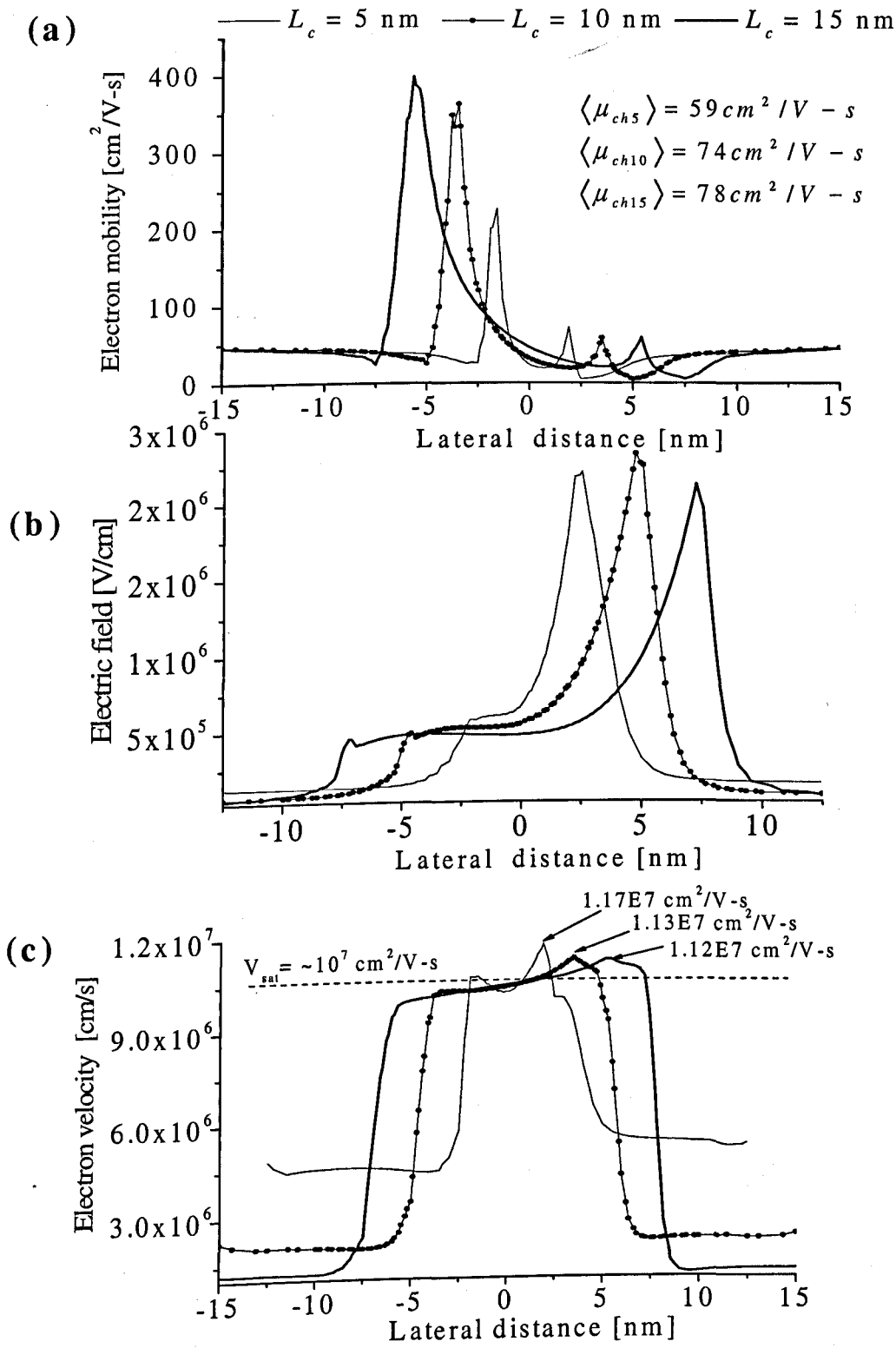


Figure 6:10 Electric field (a), electron velocity (b) and electron mobility of the selected UTBSOI MOSFETs with the channel lengths of $L_c=15,10$ and 5 nm

6.3 Conclusions

Scaling of ultra-thin body SOI MOSFETs has been investigated. In order to guarantee electrostatic integrity the silicon layer should be scaled to 3 nm, 2.5, and 2 nm in the 15 nm, 10 nm and 5nm UTB SOI MOSFETs respectively. The good performance of the scaled devices is partially due to the higher mobility compared with the scaled conventional MOSFETs of similar technology node due to low channel doping. With appropriate work functions for the gate materials the threshold voltage can be adjusted to achieve low leakage current. The main issue of scaling in UTB SOI will be the control of uniformity and thickness of the silicon layer (t_{si}).

6.4 Chapter references

- [6.1] Semiconductor Industry Association (SIA), *International Technology Roadmap for Semiconductors*, <http://public.itrs.net>, ITRS'04 update, SEMATECH, 2004
- [6.2] K. Natori "Ballistic metal-oxide-semiconductor field effect transistor", *J. Appl. Phys.*, Vol. 76 (8), pp. 4879-480, 1994
- [6.3] M. Lundstrom, "Elementary scattering theory of the Si MOSFET", *IEEE Electron Device Lett.*, Vol. 18, pp. 361-363, July 1994
- [6.4] S. Takagi, "Re-examinations of Sub-band structure Engineering in Ultra-Short Channel MOSFETs Under Ballistic Carrier Transport", *VLSI Technology Symposium*, pp. 115, 2003
- [6.5] H. Shang, H. Okorn-Schmidt, K. K. Chan, M. Copel, J. A. Ott, P. M. Kozlowski, S. E. Steen, S. A. Cordes, H.-S. P. Wong, E. C. Jones, and W. E. Haensch "High mobility p-channel germanium MOSFETs with a thin Ge oxynitride gate dielectric", *IEDM Tech. Dig.*, pp. 441-444, December 2002
- [6.6] B. Doris, "Extreme scaling with ultra-thin Si channel MOSFETs", *IEDM Tech. Dig.*, pp. 267-270, December 1982

-
- [6.7] T. Ichimori, and N. Hirashita, "Fully-depleted SOI CMOSFETs with the fully silicide source/drain structure", *IEEE Transactions on Electron Devices*, Vol. 49(12), pp. 2296-2300, 2002
- [6.8] S. D. S. Malhi, H. W. Lam, R. F. Pinizzotto, A. H. Hamdi, and F. D. McDaniel "Novel SOI CMOS design using ultra thin near intrinsic substrate" *IEDM Tech. Dig.*, pp. 107 - 110, December 1982
- [6.9] G. G. Shahidi, "SOI technology for the GHz era", *IBM J. RES & DEV.* Vol. 4 (2/3), pp.121-131, 2002
- [6.10] M. Burel, "Silicon on insulator material technology" *Electronics letters*, Vo.; 31, (14) pp. 1201-1202, 1995
- [6.11] M. Burel, "Application hydrogen ion beams to silicon on insulator material technology", *Nuclear Instrumentation and Methods in Physics Research*, B, Vol. 108, pp. 313, 1996
- [6.12] C. Mazure, G. K. Celler, C. Maleville, "Advanced SOI Substrate Manufacturing", *Semiconductor Manufacturing*, Vol.6(1), pp.34, 2005
- [6.13] Y. - K. Choi, D. Ha, T - J. King, C. Hu, "Nanoscale ultrathin body PMOSFETs with raised selective germanium source/drain" *IEEE Electron Device. Lett.*, pp. 447-448, 2001
- [6.14] J. P. Colinge, *Silicon-on-Insulator Technology: Materials to VLSI*, 3rd Edition, Kluwer Academic Publishers, 2004
- [6.15] K. M., G. Grula, J. Sleight, L. Bair, R. Stephany, R. Flatley, and P. Skerry, "A 2.0V, 0.35 μ m Partially Depleted SOI-CMOS Technology", *IEDM Tech. Dig.*, pp. 583 - 586, December 1997
- [6.16] K. Samuel, H. Fung, M. Khare, D. Schepis, W.-H. Lee, S. H. Ku, H. Park, J. Snare, B. Doris, A. Ajmera, K. P. Muller, P. Agnello, P. Gilbert, and J. Welser, "Gate length scaling accelerated to 30nm regime using ultra-thin film PD-SOI technology", *IEDM Tech Digest'01*, pp. 629-632, 2001
- [6.17] S. K. H. Fung, N. Zamdmer, P. J. Oldiges, J. Sleight, A. Mocuta, M. Sherony, S-H. Lo, R. Joshi, C. T. Chuang, I. Yang, S. Crowder, T. C. Chen, F. Assaderaghi, and G. Shahidi, "Controlling floating-body effects for 0.13 μ m and 0.10 μ m SOI CMOS", *IEDM Tech Digest'00*, pp. 231-234, 2000

-
- [6:18] Y. Taur, and T. H. Ning, *Fundamentals of Modern VLSI Technology*, Cambridge University Press, Cambridge, UK, 1998
- [6:19] S. M. Sze, *Semiconductor devices; Physics and Technology*, 2nd Edition, John Wiley & Sons, INC, USA, 2002
- [6:20] K. L. Shepard, and D.-J. Kim, "Body –Voltage Estimation of PD-SOI Circuits and its Application to Static Timing Analysis", *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, VOL. 20(7), 2001
- [6:21] D. Suh, and J. G. Fossum, "Dynamic Floating-Body Instabilities in Partially Depleted SOI CMOS Circuits", *IEDM Tech Digest 1994*, pp. 661-664, 1994
- [6:22] K. Kato and k. Taniguchi, "Numerical analysis of switching characteristics in SOI MOSFET's", *IEEE Transactions on Electron Devices* 33:1 pp. 133-139, 1986
- [6:23] W. Redman-White, R. M. Tnbroek, M. Lee, C. Edwards, M. Uren, and R. Bunyan, "Analogue design issues for SOI CMOS", *Proc. IEEE, Int. SOI Conf.*, pp. 6-8, 1996
- [6:24] Y. Omura, and K. Izumi, "Simplified analysis of body-contact effect for MOSFET/SOI" *IEEE Transaction Electron Devices*, Vol.35 (8) pp. 1391-1393 , 1988
- [6:25] J. P. Colinge, "Reduction of floating substrate effect in thin-film SOI MOSFET's", *Electron Lett.* Vol. 22, pp. 187, 1987
- [6:26] D. Esseni, E. Sangiorgi, M. Mastrapasqua, C. Fiegna, G. Celler, L. Selmi, "Ultra-Thin SOI Transistor for Ultimate CMOS Technology: Fundamental Properties and Applications", *International Journal of High speed Electronics and systems*, Vol. 12(2), pp. 333-342, 2002
- [6:27] T. Skotnicki, J. A. Hutchby, T- J. King, H. –S. Philip, and F. Boeuf, "The end of CMOS Scaling", *IEEE Circuits and Device Magazine*, January/ February, pp. 16-26, 2005
- [6:28] K. Uchida, H. Watanabe, A. Kinoshita, J. Koga, T. Numata, and S. Takagi, "Experimental study on carrier transport mechanism in ultrathin-body SOI n-and p-MOSFETs with SOI thickness less than 5 nm", *IEDM Tech. Dig.*, pp. 47-50, December 2002

-
- [6.29] T. Tezuka, S. Nakaharai, Y. Moriyama, N. Sugiyama and S. Takagi, "Selectively-formed high mobility SiGe-on-Insulator pMOSFETs with Ge-rich strained surface channels using local condensation technique" *Symposium on VLSI Tech., Digest of Technical papers* pp. 198-199, 2004
- [6.30] G. Tsutsui, M. Saitoh, and T. Hiramoto, "Superior Mobility Characteristics in (110)-Oriented Ultra Thin Body pMOSFETs with SOI Thickness Less than 6 nm", *Symposium on VLSI Tech., Digest of Technical papers*, pp. 76, 2005
- [6.31] M. Shoji, S. Horiguchi, "Electron structures and phonon-limited electron mobility of double-gate silicon-on-insulator Si inversion layers", *J. App. Phys.*, Vol. 85(5), pp. 2722-2731, 2001
- [6.32] K. Uchida *et. al.*, "Experimental Study on Carrier Transport Mechanism in Ultrathin-body SOI n-and p-MOSFETs with SOI Thickness less than 5 nm", *IEDM Tech Digest*, pp. 47, 2002
- [6.33] S. Takagi, Junji Koga, and Akira Toriumi, "Subband structure engineering for performance enhancement of Si MOSFETs" *IEDM Tech. Dig.*, pp. 219 - 222, December 1997
- [6.34] Synopsys, Medici - Two-Dimensional Device Simulation Program, Version: V-2003.12 User Guide
- [6.35] A. R. Brown, F. Adamu-Lema, A. Asenov, "Intrinsic parameter fluctuations on nanometer scale thin-body SOI devices introduced by interface roughness", *SuperLattice and Microstructures*, Vol. 34, pp. 283-291, 2003
- [6.36] F. Balestra, S. Cristoloveanu, M. Benachir, J. Brini, and T. Elewa, "Double-gate Silicon on insulator transistor with volume inversion: A new device with greatly enhanced performance", *IEEE Electron Device. Lett.*, pp. 410-412, 1987

7

Conclusion and suggestions for future work

This PhD research was aimed at achieving two major objectives. The first objective was related to the scaling of conventional MOSFETs up to the end of the ITRS roadmap. The findings regarding this first objective are drawn from investigations of five generations of ITRS transistors scaled to physical gate lengths of 35, 25, 18, 13, and 9 nm, based on a real 35 nm bulk MOSFET. The TCAD simulation tools are conventionally calibrated following the actual manufacturing process steps used by the device designers. Using this initial device structure as a reference, subsequent transistors were then carefully and methodically scaled for further investigation of performance and scaling limitations.

The second main objective of this research was the statistical investigation of intrinsic parameter fluctuations due to discrete random dopants and LER in similarly scaled conventional MOSFETs. Preliminary results of the scaling of UTB SOI FETs are also reported. Concluding remarks of the overall findings of this PhD research and some suggestions for future research programmes are presented in the following sections.

7.1 Conclusions

The scaling is based on the systematic calibration of the simulation tools and the honing of the simulation methodology with respect to the 35 nm *n*-channel and *p*-channel prototype MOSFETs. In order to achieve a realistic device structure, full scale process simulation was carried out of the scaled transistors and the reference 35 nm MOSFETs.

The output of the process simulation was used in the electrical simulations of the corresponding devices.

The calibration of both the *n*-channel and *p*-channel prototype MOSFETs was successful. The experimentally measured channel doping of the *n*-channel MOSFET was accurately reproduced by process simulation. The simulated current-voltage characteristics of the *n* and *p*-MOSFETs are in very good agreement with the experimental measurements. The agreement between the measured and the simulated device characteristics also validate the device structures obtained from the process simulation.

The subthreshold leakage current $I_{off} = 100\text{nA}/\mu\text{m}$ and the reasonably high drive current $I_{on} = 709\mu\text{A}/\mu\text{m}$ for a supply voltage of $V_{dd} = 850\text{mV}$ were well reproduced. The simulations also show that transport in the prototype device has been enhanced as a result of strain induced by a blanket deposition of Si_3N_4 used as a contact etch stopping layer.

Excellent electrostatic integrity has been achieved for the scaled devices. However, the high channel doping needed to suppress short channel effects results in significant degradation of carrier mobility and device performance. This is a real problem when it comes to the ITRS current drive requirements, I_{on} , for all scaled devices. Indeed, I_{on} cannot be achieved with the anticipated Si mobility. In fact, a mobility enhancement up to 78% is needed in the case of the 13 nm transistor. The required on-current for 9 nm transistor, $I_{on} = 2400\mu\text{A}/\mu\text{m}$, cannot be achieved using conventional MOSFETs.

Another important issue which must be taken into consideration is the solid-solubility limit of dopants in crystalline silicon. In particular, near the end of the ITRS the required active doping concentration in conventional MOSFETs (maximum In concentration in the channel of the 9 nm conventional transistor in this work is approximately $4 \times 10^{19}/\text{cm}^3$) is well above the typical solid solubility limits ($\text{In} \sim 10^{17}/\text{cm}^3$) for most dopants in silicon.

IPF in decanano MOSFETs induced by the number and position of discrete random dopants in the channel are statistically investigated, together with the impact of LER. In the case of random discrete dopants, the fluctuations of all three important MOSFET parameters namely V_T , I_{on} and I_{off} increase with decreasing transistor gate

length. For example, the random dopant induced $\pm 3\sigma V_T$ variation for the 18 nm n -channel MOSFET is $\pm 360\text{mV}$. In the 9nm device, the $\pm 3\sigma V_T$ is approximately $\pm 510\text{mV}$. This is highly unacceptable bearing in mind that the supply voltage for the corresponding technology node is expected to be $V_{dd} \sim 500\text{-}600\text{mV}$.

Similarly, the fluctuations in I_{off} and I_{on} increase with the decreasing gate length. The random dopant induced σI_{off} for the 18 nm n -channel MOSFET reaches $220\text{ nA}/\mu\text{m}$, which is 180% more than the standard deviation in the prototype 35 nm transistor. σI_{on} for the 18 nm device is $\sim 45\mu\text{A}/\mu\text{m}$.

LER, which is rapidly becoming one of the major sources of IPF has also been investigated. Different values of RMS height and correlation length have been used to investigate the impact of the LER on MOSFET parameters. The threshold voltage fluctuation in the 35 nm MOSFET as a result of LER is $\sigma V_T = 26.5\text{mV}$. This increases significantly to $\sigma V_T = 45.5\text{mV}$, in the 25 nm MOSFET, an increase of 72%. The corresponding σI_{off} for the same devices are $157\text{nA}/\mu\text{m}$ and $393\text{nA}/\mu\text{m}$ respectively. However, in the 35 nm or 25 nm MOSFET, *random discrete dopants* remain the dominant source of IPF.

Scaling of ultra-thin body SOI MOSFETs has been investigated. The main issue of scaling in UTB SOI will be the control of *uniformity* and *thickness* of the silicon layer (t_{si}). In order to guarantee electrostatic integrity the silicon layer should be scaled to 3 nm, 2.5, and 2 nm in the 15 nm, 10 nm and 5 nm UTB SOI MOSFETs respectively. Based on the simulation results of this project, electron mobility in the scaled UTB SOI transistor is about 100% higher than in the corresponding scaled conventional MOSFETs. The good performance of these scaled SOI devices is partially due to higher mobility compared with the scaled conventional MOSFETs of similar technology node due to low channel doping.

In conclusion, the results of this work confirm that after the 65 nm technology node it becomes extremely difficult to realize well-scaled conventional MOSFETs. Therefore, new device architectures or technological breakthroughs are necessary to achieve continued scaling progress. The main limitation of the scaling of transistors is the atomic structure of matter. A MOSFET cannot be smaller than an atom. The transistor post 2025 will be reaching these limitations.

7.2 *Suggestions for future research*

Suggestions for future research stemming from this work are outlined next.

1. Without abandoning conventional MOSFET structures, it may be possible to perform further investigations of scaling by introducing new materials, such as high κ gate dielectrics. However, the properties of new generations of scaled transistors with high κ materials are currently under scrutiny. This is a potentially productive research area.
2. Carrier transport in scaled devices is another area which needs further research work. This includes the introduction of new materials in the channel region with high carrier mobility (III-IV semiconductor materials) or the application of strain mechanisms in order to achieve the required I_{on} of the future generations of transistors.
3. The scalability issues of emerging device structures are the other area which needs further study and investigation. The scaling properties and limitations of non-conventional transistors, in particular UTB SOI FETs, is a challenging research opportunity.
4. The thermodynamic properties of nano-CMOS transistors and their impact on power dissipation is a major concern which needs further research work.
5. Comprehensive research on intrinsic parameter fluctuations in scaled devices, both conventional and non-conventional MOSFETs, is another area that needs further research.
6. Optimization of circuit design may be the future of intensive research work. An understanding of the properties of circuits made from scaled nano-CMOS transistors will assist the optimization process.

Appendix I

Abbreviations

$(ACF)_E$	Exponential Autocorrelation Function
$(ACF)_G$	Gaussian Autocorrelation Function
(C_A)	Mole fraction A in a compound or mixture
CMOS	Complimentary Metal Oxide Semiconductor
DADOS	Diffusion And Defects Object-oriented Simulation
DIBL	Drain Induced Barrier Lowering
FD SOI	Fully Depleted Semiconductor On Insulator
IPF	Intrinsic Parameter Fluctuations
ITRS	International Technology Roadmap for semiconductors
LER	Line Edge Roughness
LSS	Lindhard Scharff Schoitt (LSS ion stopping theory)
LWR	Line Width Roughness
MOSFET	Metal Oxide Semiconductor Field Effect Transistors
MPU	Micro Processing Unit
PD SOI	Partially Depleted Semiconductor On Insulator
RMS	Root Mean Value
RTA	Rapid Thermal Annealing
SCE	Short Channel Effect
SOI	Silicon On Insulator
SSRC	Supper Steep Retrograde Channel
STI	Shallow Trench Isolations
TCAD	Technology Computer Aided Design
TEM	Transmission Electron Microscopy
UTB SOI	Ultra Thin Body Semiconductor On Insulator

Appendix II

Maximum solid solubility of selected atoms in crystal silicon[†]

Symbol	Name	Concentration [atoms/cm ³]	Temperature [°C]
Al	Aluminium	2.0×10 ¹⁹	1200
As	Arsenic	1.8×10 ²¹	1150
B	Boron	1.0×10 ²⁰	1000
C	Carbon	3.5×10 ¹⁷	-
Co	Cobalt	1.0×10 ¹⁶	1200
Ga	Germanium	3.2×10 ¹⁹	1100
In	Indium [*]	8.0×10 ¹⁶	1100
Li	Lithium	7.0×10 ¹⁹	1200
N	Nitrogen	~6.0×10 ¹⁸	-
O	Oxygen	1.2-1.4×10 ¹⁸	-
P	phosphorus	3.0×10 ²⁰	850
Sb	Antimony	4.65×10 ¹⁹	1066
Ti	Titanium	1.0×10 ¹⁴	1200

^{*} By using laser annealing, maximum solid solubility of indium in crystal silicon (1.5×10¹⁸/cm³) might be achieved

[†] Source: Ref: [4:30]

Appendix III

Publications and conference papers

G. Roy, F. Adamu-Lema, A. R. Brown, S. Roy and A. Asenov, "Simulation of Combined Sources of Intrinsic Parameter Fluctuations in a 'Real' 35 nm MOSFET", *Proc. European Solid-State Device Research Conference (ESSDERC)*, 12-16 September, Grenoble, France, pp.337-340 (2005)

B. Cheng, S. Roy, G. Roy, F. Adamu-Lema, A. Asenov, "Impact of Intrinsic Parameter Fluctuations in Decanano MOSFETs on yield and functionality of SRAM Cells", *Solid-State Electronics*, Vol.49, pp. 740-746, 2005

L. Yang, J. R. Watling, F. Adam-Lema, A. Asenov, J. R. Barker; "Scaling study of Si and strained Si n-MOSFETs with different high-k gate stacks", *IEDM Tech. Dig.*, pp. 597 - 600, December 2004.

F. Adamu-Lema, G. Roy, A. R. Brown, A. Asenov and S. Roy, "Intrinsic Parameter Fluctuations in Conventional MOSFETs at the Scaling Limit: A Statistical Study", *Journal of Computational Electronics*, Vol.3, pp.203-206 (2004)

L. Yang, J. R. Watling, F. Adam-Lema, A. Asenov, J. R. Barker "Simulations of Sub-100 nm Strained Si MOSFETs with High- κ Gate Stacks" *Journal of Computational Electronics*, Vol.3, pp.171-175 (2004)

L. Yang, J. R. Watling*, R. C. W. Wilkins, F. Adamu-Lema, A. Asenov, and J. R. Barker, "Interface Roughness Scattering in sub-100nm Si and Strained Si n-MOSFETs: A Monte Carlo Study", Submitted to journal of *IEEE Transaction on Electron Devices*

A. R. Brown, F. Adamu-Lema and A. Asenov, "Intrinsic Parameter Fluctuations in Nanometre Scale Thin-body SOI Devices Introduced by Interface Roughness", *Superlattices and Microstructures*, Vol.34, pp.283-291, (2003).

A. R. Brown, F. Adamu-Lema and A. Asenov, "Intrinsic Parameter Fluctuations in UTB MOSFETs Induced by Body Thickness Variations", *Silicon Nanoelectronics Workshop*, 8-9 June, Kyoto, Japan (2003)

A. Asenov, M. Jaraiz, S. Roy, G. Roy, F. Adamu-Lema, A. R. Brown, V. Moroz and R. Gafiteanu, "Integrated Atomistic Process and Device Simulation of Decananometre MOSFETs", *Proc. International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, IEEE Cat. No. 02TH8621, pp.87-90 (2002)

